METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from NIDCR, Grant Numbers: DE11601 and DE12920.

FIELD OF THE INVENTION

The present invention relates to the fields of genetic screening and molecular biology. More specifically, the invention provides compositions and methods that may be used to advantage to isolate and detect a palmoplantar keratoderma predisposing gene, cathepsin C (CTSC), some mutant alleles of which cause susceptibility to certain pathological disorders, in particular Papillon-LeFevre Syndrome, Haim-Munk Syndrome and certain forms of early onset periodontal diseases. More specifically, the invention relates to germline mutations and functional polymorphisms in the CTSC gene and their use in the diagnosis of predisposition to palmoplantar ectodermal disorders/ dysplasias and periodontal diseases. The invention also relates to the therapy of palmoplantar ectodermal disorders/dysplasias and periodontal diseases which have a mutation or functional polymorphisms in the CTSC gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for treating and alleviating disease symptoms. Finally, the invention relates to the screening of the CTSC gene for disease-related mutations, which are useful for diagnosing the predisposition to additional disorders and dysplasias, including but not limited to prepubertal periodontitis, early onset periodontal disease or other forms of gum disease.

BACKGROUND OF THE INVENTION

Various publications or patents may be referenced in this application by numerals in parentheses to describe the state of the art to which the invention pertains. Full citations for these references are provided at the end of the specification. Each of these publications or patents is incorporated by reference herein.

Most forms of inflammatory periodontal disease can be successfully treated and managed. As a result, the ultimate goal of periodontal therapy has changed from that of simply arresting disease progression to one aimed at regenerating the supporting tissues. Unfortunately, not all forms of periodontal disease respond to treatment. Severe periodontitis that is resistant to conventional periodontal treatment has been recognized in a number of monogenic conditions. Certainly some of the most intriguing and dentally challenging of these conditions include Papillon-Lefevre syndrome (PLS), Haim-Munk syndrome (HMS) and periodontal diseases.

In 1924, Papillon and Lefevre described two siblings, the products of a first cousin mating, with a condition characterized by diffuse transgradient palmoplantar keratosis (PPK) and the premature loss of both the decidous and permanent dentitions. This condition came to be known as Papillon-Lefevre syndrome and subsequently over 200 cases have been described. The hallmarks of PLS are palmoplantar keratosis and rapid periodontal destruction of both dentitions. An increased susceptibility to infection has been reported in approximately 20% of PLS patients. Additional findings include intracranial calcifications, retardation of the somatic development, follicular hyperkeratosis and onychogryphosis. Clinical findings reported in PLS patients
suggest that the clinical expression of the condition is

highly variable. Unfortunately, to date, no pathognomonic disease marker exists allowing definitive diagnosis of PLS.

In 1965, Haim and Munk described an unusual syndrome in four siblings of a Jewish religious isolate from Cochin, India [21]. In addition to congenital palmoplantar keratosis and progressive early onset periodontal destruction, other clinical findings shared by these individuals included recurrent pyogenic skin infections, acroosteolysis, atrophic changes of the nails, arachnodactyly, and a peculiar radiographic deformity of the fingers consisting of tapered pointed phalangeal ends and a clawlike volar curve. Subsequently pes planus was reported to be associated with the syndrome [24]. This was the first reported association of these clinical findings, and the condition became known as Haim Munk syndrome, or keratosis palmoplantaris with periodontopathia and onychogryposis (HMS; MIM245010)[22]. Although the palmoplantar findings and severe periodontitis were suggestive of the Papillon-Lefevre syndrome (PLS; MIM245000)[3], the association of other clinical features, particularly nail deformities and arachnodactyly, argued that HMS was a distinct disorder.

PLS and HMS are classified as type IV palmoplantar ectodermal keratodermas [2]. The unique presence of severe, early onset periodontitis distinguishes PLS and HMS from other PPKs and raises the question of whether they result from the variable clinical expression of a common gene mutation, are allelic mutations at the same genetic locus, or result from expression of gene mutations at separate loci. Although Haim and Munk's initial report proposed HMS was a distinct entity, Hacham-Zadeh and co-workers referred to the disorder as Papillon-Lefevre syndrome and cited Gorlin's suggestion that HMS was a clinical variant of PLS [23,25]. In his review of PLS cases reported in the literature, Haenke

[5] summarizes an extensive list of clinical findings reported in PLS affected individuals, including increased susceptibility to infections, ectopic cranial calcifications and nail anomalies [5,26]. It is unclear if these additional clinical features are coincidental findings that may be segregating in a particular family or if they are etiologically related to a syndrome with a very variable clinical expression. Because PLS is an uncommon condition, and generally occurs only in a single generation it is difficult to determine if these occasional reports of associated clinical findings are etiologically related to PLS. Additionally, consanguinity is common among parents of PLS cases and therefore, it may be expected that an increased number of rare recessive conditions may be seen. Such is likely the case for the reports of mental retardation associated with PLS [5].

Pre-pubertal periodontitis (PPP) is a rare and rapidly progressive disease that results in destruction of the periodontal support of the primary dentition. The condition may be localized (usually to deciduous molars) or generalized. The localized form begins at approximately 4 years of age and is associated with only mild gingival inflammation in the presence of relatively little plaque. The generalized form begins earlier, immediately after eruption of the deciduous teeth. It is associated with severe gingival inflammation and hyperplasia, although significant gingival recession has also been described as an associated clinical feature. The attachment loss appears to be continuous rather than intermittent as with most other forms of periodontitis.

A varied clinical phenotype has been reported for PPP, probably reflecting the fact that the term PP describes an etiologically heterogeneous group of conditions that share an overlapping clinical presentation. Although PPP can occur as an isolated finding, many reports of PPP describe an increased

systemic susceptibility to infections. Children with PPP have frequent inner ear infections and infections of the respiratory tract [39,40]. Prepubertal periodontitis is known to be associated with Papillon Lefevre syndrome and with a number of systemic disease states that share an increased susceptibility to microbial infections.

To date, no pathognomonic disease marker exists for most PPKs allowing for definitive diagnosis. The present invention provides such a disease marker and methods of use thereof having diagnostic and prognostic utilities for several PPKs and many periodontal diseases.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods which allow for genetic screening and diagnosis of certain palmoplantar keratodermas and periodontal disease states in affected individuals. In accordance with the present invention, it has been discovered that mutations or functional polymorphisms in the cathepsin C gene (CTSC) give rise to certain pathological conditions including PLS, HMS and periodontal diseases. Mutations or functional polymorphisms associated with the disease state are those which give rise to a altered, truncated, misfolded or otherwise non-functional CTSC polypeptides. Polymorphisms in the CTSC sequence which do not affect the nature of the encoded protein are not associated with PLS, HMS or periodontal disease.

Thus, in one embodiment of the invention, a method is provided for determining the presence of alterations in CTSC encoding nucleic acids which give rise to altered CTSC proteins. The wild-type CTSC nucleic acid sequence and its encoded amino acid sequence are known. See SEQ ID NOS: 1-3 provided herein. This sequence information facilitates the identification of genetic changes that give rise to aberrant CTSC proteins.

CTSC mutations specifically associated with PLS,

HMS and PPP are described herein and are set forth in Table 1. Accordingly in one embodiment of the invention, nucleic acid molecules encoding altered CTSC proteins are considered to be within the scope of the present invention. In a preferred embodiment of the invention, the altered CTSC nucleic acid has at least one of the alterations set forth in Table 1.

In a further embodiment of the invention, nucleic acid probes which specifically hybridize to the human altered CTSC-encoding nucleic acids described herein and not to wild-type CTSC encoding nucleic acids are provided. In a preferred embodiment, the probes specifically hybridize with altered CTSC encoding nucleic acids having at least one of the alterations set forth in Table 1.

In yet another embodiment of the invention, a mutated CTSC protein encoded by the altered CTSC encoding nucleic acids of the invention is provided. Preferably such CTSC proteins are encoded by a nucleic acid containing a mutation as set forth in Table 1. Also provided are assays for biochemically assessing altered cathepsin C activity. Antibodies immunologically specific for altered CTSC proteins are also contemplated to be within the scope of the present invention.

In another aspect of the invention, a method for detecting a germline alteration in a CTSC gene is provided. In a preferred embodiment the alteration is selected from the group consisting of the alterations set forth in Table 1. The method comprises analyzing a sequence of a CTSC gene or CTSC RNA from a human sample or analyzing a sequence of CTSC cDNA made from mRNA from a human sample and comparing sequences so isolated to the wild type sequence encoding CTSC. Inasmuch as certain alterations of the CTSC coding sequence may not alter the function of CTSC, methods are provided for assessing the enzymatic activity of proteins encoded by nucleic acid molecules which do not possess the wild

type CTSC sequence.

In yet another embodiment of the invention, kits are provided for detecting the presence of an altered CTSC encoding nucleic acids in a biological sample. An exemplary kit comprises the following: i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1; ii) reaction buffer; and iii) an instruction sheet. Kits for detecting the presence an altered CTSC proteins in a biological sample are also provided. Exemplary kits for this purpose comprise: i) antibodies immunologically specific for the altered CTSC proteins of the invention; ii) a solid support with immobilized CTSC antigens as a positive control; and iii) an instruction sheet. Optionally, anti-CTSC antibodies used for this purpose may contain a detectable label or tag for used in isolating or detecting immune complexes.

Various terms relating to the biological molecules and cells of the present invention are used throughout the specifications and claims.

With reference to nucleic acids used in the invention, the term "isolated nucleic acid" is sometimes employed. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule. An isolated nucleic acid molecule inserted into a vector is also sometimes referred to herein as a Arecombinant@ nucleic acid molecule.

With respect to RNA molecules, the term "isolated

nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

With respect to antibodies, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization

under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence. Appropriate conditions enabling specific hybridization of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is set forth below (Sambrook et al., 1989):

$T_m = 81.5°C + 16.6\text{Log}\ [Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ \text{formamide}) - 600/\#bp\ \text{in duplex}$

As an illustration of the above formula, using [Na+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

The term "promoter region" refers to the transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. In the present invention, the use of SV40, TK, Albumin, SP6, T7 gene promoters, among others, is contemplated. Specific promoters for the yeast and mammalian expression systems of the invention are available and known to those of ordinary skill in the art.

The term "operably linked" means that the

regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement of transcription units and other transcription control elements (e.g. enhancers) in an expression vector.

The phrase "functional polymorphism" refers to a change in wild type CTSC coding sequence giving rise to altered cathepsin C activity as assayed using conventional methods.

The term "oligonucleotide," as used herein refers to primers and probes of the present invention, and is defined as a nucleic acid molecule comprised of two or more ribo- or deoxyribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on various factors and on the particular application and use of the oligonucleotide.

The term "probe" as used herein refers to an oligonucleotide, polynucleotide or nucleic acid, either RNA or DNA, whether occurring naturally as in a purified restriction enzyme digest or produced synthetically, which is capable of annealing with or specifically hybridizing to a nucleic acid with sequences complementary to the probe. A probe may be either single-stranded or double-stranded. The exact length of the probe will depend upon many factors, including temperature, source of probe and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide probe typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The probes herein are selected to be complementary to different strands of a particular target nucleic acid sequence. This means that the probes must be sufficiently complementary so as to be able to "specifically hybridize" or anneal with their respective target

strands under a set of pre-determined conditions. Therefore, the probe sequence need not reflect the exact complementary sequence of the target. For example, a non-complementary nucleotide fragment may be attached to the 5' or 3' end of the probe, with the remainder of the probe sequence being complementary to the target strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the sequence of the target nucleic acid to anneal therewith specifically.

The term "primer" as used herein refers to an oligonucleotide, either RNA or DNA, either single-stranded or double-stranded, either derived from a biological system, generated by restriction enzyme digestion, or produced synthetically which, when placed in the proper environment, is able to functionally act as an initiator of template-dependent nucleic acid synthesis. When presented with an appropriate nucleic acid template, suitable nucleoside triphosphate precursors of nucleic acids, a polymerase enzyme, suitable cofactors and conditions such as a suitable temperature and pH, the primer may be extended at its 3' terminus by the addition of nucleotides by the action of a polymerase or similar activity to yield an primer extension product. The primer may vary in length depending on the particular conditions and requirement of the application. For example, in diagnostic applications, the oligonucleotide primer is typically 15-25 or more nucleotides in length. The primer must be of sufficient complementarity to the desired template to prime the synthesis of the desired extension product, that is, to be able anneal with the desired template strand in a manner sufficient to provide the 3' hydroxyl moiety of the primer in appropriate juxtaposition for use in the initiation of synthesis by a polymerase or similar enzyme. It is not required that the primer

sequence represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary bases may be interspersed within the oligonucleotide primer sequence, provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the extension product.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1E are a series of clinical photographs showing palmoplantar keratosis and periodontal disease in PLS study patient. Fig. 1A: palmar hyperkeratotic lesions; Fig. 1B: plantar hyperkeratotic lesions; Fig. 1C: hyperkeratotic lesions affecting the knees; Fig. 1D: periodontitis involving erupting permanent dentition and Fig. 1E: periapical radiographs showing severe alveolar bone loss affecting erupting permanent teeth.

Figure 2 shows haplotype data for chromosome 11q short tandem repeat polymorphisms (STRP) markers spanning the PLS gene locus. Segments which are likely to be homozygous by descent are boxed. Arrows indicate recombinant events. Individuals 7 and 22 share a common haplotype for D11S1979, D11S1887, D11S1780, D11S1367, D11S931, and D11S4175.

Figure 3 depicts pedigree and sequence analysis of *CTSC* exon 6. The numbering of the wildtype sequence shown above the figure is based upon the genomic sequence of *CTSC*. See SEQ ID NO: 1. Circles represent females and squares represent males. Filled symbols indicate affected individuals. Half-shading indicates carriers based upon DNA sequencing results. All affected individuals are homozygous for the specific CTSC

mutations. Arrows indicate the position of the mutation. This family has a nonsense mutation (856 C->T) at codon 286 resulting in a truncated protein of 286 amino acids.

Figures 4A - 4D show pedigrees and sequence analysis of *CTSC* exon 7 for 4 Families with PLS. Symbols are as described for Figure 3. Fig. 4A: Family with a single base pair deletion (1047delA) of *CTSC* resulting in a frameshift and premature termination. Fig. 4B: Family with a 2bp deletion (1028-1029delCT) of *CTSC* resulting in a frameshift and premature termination. Fig. 4C. and Fig. 4D: pedigrees of families with a nonsense mutation (1286G->A) at codon 429 resulting in a truncated protein of 428 amino acids. The father in family C is deceased and no sample was available for analysis.

Figures 5A-5C is a schematic diagram of the CTSC gene showing the locations of the mutations described herein. Panel 5A. Genomic structure of *CTSC* gene with introns shown as solid lines and exons depicted as boxes. The 5' and 3' untranslated regions are shown as filled boxes. Panel 5B. Coding region of *CTSC* gene. The amino acid numbers are shown at the end of each exon. Mutations listed in Table 1 are shown according to their genomic locations with Missense, *; Nonsense, ♦; Insertion, ■; and Deletion, □. The splicing site mutation is indicated by an arrow. Panel 5C. Subunit structure of *CTSC* polypeptide with SP, signal peptide; P1, 13.5 kDa pro-region; P2, 10 kDa pro-region; H, heavy chain; and L, light chain. The 10 kDa pro-region is cleaved out upon activation. The disulfide bond within the 13.5 pro-region is shown. The glycosylation sites are indicated by filled circle and arrows indicate the active sites.

Figures 6A-6F show a series of micrographs

depicting the clinical and radiographic findings in Haim Munk syndrome. Fig. 6A: dermal involvement of fingers in individual #34. Fig. 6B: Individual #34 radiograph of terminal phalanges of the fingers showing marked thinning increasing towards the distal, tapering pointed ends showing a claw-like volar bend. Fig. 6C: Individual #17 palmar keratosis; Fig. 6D: Individual #17 plantar keratosis: Fig. 6E: Individual #17 gingival inflammation; Fig. 6F: Individual #17 radiograph showing alveolar bone destruction associated with gingival inflammation shown in 6D.

Figure 7A shows pedigree of Cochin descendents segregating Haim Munk syndrome (HMS). Numbered individuals have been analyzed for the current study. Circles = females, squares = males, shaded symbols = HMS affected individuals. Double lines = consanguinity. Individuals #10,11 * = second cousins. Numbers inside circles, squares and diamonds indicate the number of additional offspring not examined in this study. Sibships described in previous reports are indicated and referenced below the pedigrees. The subjects of Haim and Munks original report (1965) are individuals 33, 34, 35, and 36. Half-shading indicates carriers based upon DNA sequencing and/or restriction enzyme analysis. Unshaded numbered individuals represent non-carriers based upon DNA analysis. Figure 7B shows a pedigree of a Turkish family segregating PLS. Numbered individuals were available for study. Half-shading indicates carrier based upon DNA analysis. Individual 77 is a non-carrier based upon DNA sequencing.

Figures 8A and 8B show the results of sequence analysis of exon 6 of CTSC. The numbering of the wildtype sequence is based upon the cDNA sequence of CTSC. See SEQ ID NO: 1. Fig. 8A: Family A (Cochin isolate diagnosed with Haim Munk syndrome) from Figure

1. Affected individuals are homozygous for a 857A->G missense mutation which results in a conserved glutamine being changed to an arginine (Q286R). Representative sequences are shown for individuals #36 (affected) and #31(carrier). Fig. 8B. Family B from Figure 1. Affected individuals are homozygous for a 856C->T nonsense mutation which results in a premature stop codon at position 286 (Q286X). The Q286X mutation has been previously reported in an inbred Turkish family [12].

Figure 9 depicts a gel showing the results of restriction enzyme analysis of Q286R mutation in Family A of Fig. 7. A 465 bp fragment of exon 6 was amplified and subjected to restriction digestion with AvaI as described under methods. The Q286R mutation introduces a new AvaI site. After digestion and electrophoresis through 1.8% agarose gels, wildtype individuals exhibit bands of 465 bp, affected individuals have bands of 404 and 61 bp, and carriers have bands of 465, 404, and 61 bp. M. 1 kb ladder (Gibco). Lane 1. Individual #5 uncut, demonstrating 465bp amplicon. Lane 2. Individual #5 cut with AvaI. Only the 465 bp fragment is observed. Thus individual #5 has the wildtype sequence on both alleles. Lane 3. Individual #31 uncut. Lane 4. Individual # 31 cut with AvaI. The 465 and 404bp fragments are visible, confirming that individual #31 is a carrier of the Q286R, consistent with the sequencing results shown in Figure 3A. Lane 5. Individual #34 uncut. Lane 6. Individual #34 cut with AvaI. The 404 and 61bp fragments are indicated by arrows.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect mutated forms of the lysosomal

protease cathepsin C (CTSC) gene, associated with autosomal recessive disorders characterized by palmar hyperkeratosis and/or periodontitis. More specifically, the present invention relates to germline mutations in the CTSC gene and their use in the diagnosis of predisposition to such pathological conditions. Additionally, the invention relates to germline mutations in the CTSC gene in other palmoplantar ectodermal disorders and dysplasias and their use in the diagnosis and prognosis of such pathological conditions. The invention also relates to the therapy of palmoplantar ectodermal disorders and dysplasias which have a mutation or functional polymorphism in the CTSC gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs which may have therapeutic value. Biochemical assays are provided for the assessment of altered activity of aberrant CTSC enzymes encoded by the mutated CTSC encoding nucleic acids of the invention. Finally, the invention relates to the screening of the CTSC gene for mutations, which are useful for diagnosing the predisposition to ectodermal disorders and dysplasias.

The present invention provides an isolated polynucleotide comprising all, or a portion of the CTSC locus or of a mutated CTSC locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the CTSC locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the CTSC locus, and may further include a

step of providing a set of polynucleotides which are primers for amplification of said portion of the CTSC locus. The method is useful for either diagnosis of the predisposition to PPKs or the diagnosis or prognosis of keratodermal disorders/dysplasias and periodontal diseases.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the altered CTSC locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the CTSC locus, the kits comprising a polynucleotide complementary to the portion of the CTSC locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the CTSC locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the CTSC locus.

The present invention further provides methods of screening the CTSC gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the CTSC locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the CTSC locus. Exemplary primers are set forth in Table A.

The method is useful for identifying mutations for use in either diagnosis of the predisposition to keratodermal disorders/dysplasias and periodontal diseases or the diagnosis of such disorders.

The present invention further provides methods of screening suspected CTSC mutant alleles and functional

polymorphisms to identify mutations in the CTSC gene. In addition, the present invention provides methods of screening drugs for therapy and to identify suitable drugs for restoring CTSC gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at aberrant cells associated with keratodermal disorders and dysplasias. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the CTSC locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the CTSC protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of CTSC. These may functionally replace the activity of CTSC in vivo.

It is a discovery of the present invention that mutations in the CTSC locus in the germline are indicative of a predisposition to keratodermal disorders/dysplasias and periodontal diseases. The mutational events of the CTSC locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

A major gene locus associated with the keratodermal disorders and dysplasias of the invention has been localized to a 2.8 cM interval on chromosome 11q14 of the human genome. This region contains a genetic locus, CTSC. The CTSC message is expressed at high levels in a variety of immune cells including polymorphonuclear leukocytes, macrophages and their precursors. This gene is expressed in the palms, soles, knees, and oral keratinized gingiva.

The CTSC gene was originally reported to consist of 2 exons. US Provisional Application 60/165,016 from which the present application claims priority, describes mutations in Exons 1 and 2. The mutations described actually fall within Exons 6 and 7. Reference numerals

to the altered amino acids are the same as those in US Provisional 60/165,016, only the nucleotide numbering has changed to reflect the actual genomic structure of the CTSC gene which is now known to contain 7 exons. The sequence encoding the wild type human CTSC gene is provided below (SEQ ID NO: 1):

AGGGAGATAT AAGTGAATAA TTTGGACCTG CTCTCTTTGA ATGTTTATAA TCTGGTGGAA
AAAAAATGGA CATATGAATA TTGATTTGTG ACCAGTGCAA AGGGGGCAAA AATTCATATC
CCAAAGAAAA CGGGGACACA TCAGGTCTGT CTTGTTCATC ACTGTGTCCA CAGGGCCTGA
CACCTAGTAG GCTCAGTGGG AGAAAGGAGC CCCAATTACC AACAAAAGCC AGGAAAGAAC
GGGAGGCTCT TACGGAAAAG GGTGATACTT AAACTGAGCA AGGAGGCACC TGGAAATAGT
GCCACCTAAT AATTTTTGGC GATCAGACTG ACACACTAGA ACGGTTCATA AGACCAGCCT
TCTCCCATTG GCTAGCTTCC TTCCTCACCC TTCTCACCCT GGGCAAGCCG CTTCCTCTCT
CTGGGCCTCT TGCTTTTCCT CTGTAACATA AAAGGGGTTG AGCAATATCA TCTCTGAGAG
CGCCATGTGT GTGCGTGCCA GAGGGAAAAC CCCCACAACG CTAATACATC AAAACTGCAG
GTTTGCACAA AAACTGAATT CTGCTGAATG CAAACAGGCA AACAGCATTT ACCAGGAAAC
AAAACAAAAT CAAGCACATA AAAAAGTAGG AAGAGTTGGA AAACGGAAGG AAGATAAGTT
CTCAAACAGC TGGAATAGTT GATGTTAGCT AGCGAAGTTT TTCAGAGGAA AAAACAAGAA
GTTGGTTATG AGGCAAGTGG ACCTGAGAAA AAAGACTAAA GGGGAAGAAT AGCAAGTAAA
ACAGAACTCC ACTTGCTAGA TCTCTCCCTC TGTCGCGCTC TTTCACCTGA CCCACTCCCT
TATTCCCCCC ACACCCTTTC CTTCTCTCCC TACGTTACCG CACAGGAACG AAGTCTGGGT
CATGTGCGGA CCGCTTGTGG CTCTTAAATC CTCTTTTTGT CACCCTGGCC GTGCAAAATT
TTGAAACGTC CCTCGGCAAA AAAAATAAAA ATAAAAAAAA AAAATCTGTC CCTGGCCTCT
TCCCTAGTTC TGGGTCCAGT TGCAGCCAAG TGAGGGGCAG CGCGCGCTCC CAAGTCCCCG
TTTCAGAGAC GCGCACGCGC CTGGCGCCCA ACCCCCAATC CCCTGCTGCT CAGTGACCCC
GCCCACGGGT TTCCGGGCCG GCGTAGCTAT TTCAAGGCGC GCGCCTCGTG GTGGACTCAC
CGCTAGCCCG CAGCGCTCGG CTTCCTGGTA ATTCTTCACC TCTTTTCTCA GCTCCCTGCA
GCATGGGTGC TGGGCCCTCC TTGCTGCTCG CCGCCCTCCT GCTGCTTCTC TCCGGCGACG
GCGCCGTGCG CTGCGACACA CCTGCCAACT GCACCTATCT TGACCTGCTG GGCACCTGGG
TCTTCCAGGT GGGCTCCAGC GGTTCCCAGC GCGATGTCAA CTGCTCGGTT ATGGGTAAGC
CGCCGGCTCG GCAGTCCTCC GGGTCGTCCT TTCTGCCCTT GAGCCCCTAA CGCAGCGCCA
CGCCAACTAC CGCTTCCCCC CAGGCAGACG CTTGTGGGTG GCCAGAGCAT CTTGACTGGA
TTCGGGGACC CTTGGGGACC TTCTTCCCCG CCAGGCTCGC GAAGTTAAAG TTCATCTGCT
GAGAACTTCT AACTCCACAC TTTCTTGGTT ATCTTGGGGA CTCAACACTT TGATCAAGAA
CTTTTTTATT CCTCCCGCTT AATTTTGTTT GCTTTGAGAG AGACTTGGGA ACTGCAATCG
TTTGGTTCTC CAGTCCGATC TGGTAGCGTT ATTTTTAAAA TTTATTTTTA TTTTTTATTA
CTATTTTACT AGTGAAGATA GATGAGCTCA GAGACTCTCG AGGATATAGC ATGAAGTTTT
CTCTTTTTGT TAGATGGTGG GAAAGGGACT TTCTGCCCAG CGATTTTGGT TTGAGCGGGT
GTTGATGAGT ACTAGAAAAC GGCTAGTACC ACTCTGCATT GTTTCATGCA TTGCAAGGAG
GTAAAAATTT TTTAAAAAAT TAATAACAAA GAAAACTTAA CTCTGAACCT AGTAATTAGA
AATGCCCAGA GTCTGCACAA TGTTTGGCTC ATGGAAGGCT CTCAATAAAT ACCTAGTGTT
TGAACATACT GGAGATATTC CATATGCCTT CAGATAACAT GGTTACCCCT AGAACAAAGA

ACCTAGTGAA GGGGGTGGGG TGGAAAAAAG ACAATCACTA GTTAGAAGAG TCACACTGTG
GTCTACCCAA ACCCTCTTAC AGCCTGTGAC TTCTGCAGAG TCAGTAAAAA ATCAGCTATA
ATTTTCTTGT CAACAGAACA AAGATTAAGT CATTTGTTAC TAAAGAAATA CCTTTTTAAC
TGACTGTATA GCATTTCATA ATCCTGAACA TTGTAACTTT TTTTTTTTTT TTTTTGAGAA
TGGAGTCTTG CTCTGTCACC CAGGCTGTAG TTCAGTGGCG GGATCTCGGC TCACTGCAAC
CTCCGCCTCC CGGGTTTAAG CGATTCTCCT GCCTCAGTCT CCCGAGTAGC TAGGATTACG
GGCACATGCC ACCACGCCCG GCTAATATTT TGTATTTTTA GTAGAGACGG GTTTCACTGT
GTTAGCCAGG ATGGTCTCCA TCTCCTGACC TCGTGATCCG CCCGCCTCGG CCTCCCAAAG
TACTGGGATT ACAGGCGGGA GTCACCACGC CTGGCCTTTT TTTTTTTTTT TTTTTTAAAT
TTGAGTTTGA AGGTTGGGGC AGAAAGAGAT CAGGATTTGC ACTGCCCTGT CACATGCAAT
CTCCCATGTC AGAGCATTAC TCTCAAACAT GGAAAAACTT TAAAATACAC AACTCTCCAG
ATGGACACAG CTGAATCATT TTCAGGAAGG CTCTGCCTAT AAAATTACCT TTGGCCTTAA
TTCAGTAATT AAAGGCACCC ACAGGTCTGA GCCCTCATCT ATGAAGGTTA AACGTATCAT
TCCTATCAGC ACTAATTGGT TTTATAGGAT ACAGACCTCT AGTTTGCTGA ATAAACTTTG
GAAGGATTCA GATTATGAGG TTCTAAACTG TAGAGCCTTT GAGGAGAAAG GTACCCCATT
TCTTCTCCGA ATAGATCATT TTGTGTCTTC TCCTGGGCTT AGCACATTGT CTTCCTTAGG
ATCTGATAGT CTGTTTATTT ATCTTTTTGT TGTCATACCT TTTATTCTTG CATTTCCCAC
TTTTGACTAG CATTTTGCCT TTTCTCCGTT TCTGGAAGCC TGTAATTTTC AACATTCCCC
ATTTCTCCTT TTGTTCAGTG GAAAAATTTC TTCAGTGTTG GGATGCTCTG GAAAGGTACT
GTAGTTTTGG GGGTCTCCCT GCCTGCTGGG CTGCATAGCA TATCTCTCTT TTTTGAGACA
AGGTCTCACT GTGTTGCCCA GGTTGGAATG CAGTTGTGCA ATCATAGCTC ACTGCAGCCT
CCAACTCCTG GGCTCAAGCA ATCCTTCCAC CTCAGCTTCA GCCGCTCAAA TAGCTGGGAC
TACAGGCAAT GCACCACCAT AACCAGCGAA TTTTTAAAAA TTATTTTTGT AGAGACAGGG
TCTCACTATG TTGCCCAAAC TAGTCTTGAA GTTCTGACCT CAGCCCCGCA AAGTGCTAGG
ATTACAGGCG TGAGCCACTG CTTCTGCTTG CATTATCTCA AATTTTTAGA GCCTAACTTC
ACAGTCTTGC CTCGGAGCAG GAACTGTTTG AGGCAACAAG ATGGAGCCTC TGGTTTCTTC
ACTAGGCAGA CTGTGCTCAA ACTGGGTAGC ATGAAAGGAA AGCAGCACAA TTAAAATTGA
AATTGGGGGA TTCTTTCCCC TATTGAAATC CAAATAATTT TCCTTTATTG TGCTTTTTTT
TTTTTTCTTT AGGACCACAA GAAAAAAAAG TAGTGGTGTA CCTTCAGAAG CTGGATACAG
CATATGATGA CCTTGGCAAT TCTGGCCATT TCACCATCAT TTACAACCAA GGCTTTGAGA
TTGTGTTGAA TGACTACAAG TGGTTTGCCT TTTTTAAGGT TAGTTTTGTT GGAAGTTGGA
TTTACATTTT CAATGATTTG ATATCTGAAA CCTCTTCTGA TTAGTAGACC CTCAGAATTT
TAATTTTAGA TTAAGAAGAT GACCGGAATT GACACCACTC TTCCCAAGAG TAGTAGTAGT
TGGTATAATT TTGCCACTTT ATTTAAAATT ATGATTTTAG TAGGTTTACA AAATACCAGT
GCTACATTTG AATATGTATA AATTATGTTT AAAATTTACA TTTTGGTAAG TATGACTAAA
TTCTTAATTT ATTTTCCTTA TTACTACCAC TTTATTTCTA AATGTTGCCA TAGTCATTTG
GCTTTGTTCT AAATCTGTAG GAAAGATAGA GAGATTACAC ATTTTGTTTT CTTGCAGTTA
CTATGCTGTC CTTCCTATCA CTACCTGTTG GCTGAGGTAG TGATAGGCCT AAATGATTCA
TTATCTTAAA TGTACTAAAT ATGTTGAGTA ATTTTTTCTT CTAAACTAAC AGAAAGAGAG
AACCTAGGAG TTACTCCCTT AGGCTGGTTA AAGTGAAAGG TAGCCAAGTC AACCCAGCTT
GTTTCCTTCT CTCATTAGGA AAGAACTATT GTTCATTCTC ATAACACACT TTTTCCAATT
GCAAACATAC TCAGGGTTAA AATAGTTTAG CACAAATTGC AGCCCATTTC ATTTGTTCTT
CACAAGCTGG AACTTTTCTT GTAAGCTAAA TATTAAATGG TTCAAGTAAA TTGGATACAT
AAGCCTGAAA CTAGGCGTTT CTCATTATAC ATAGAGTATA AATTAAGACA GACTTTTTCA

TGGTGAAAGG TTTACAGCCT TTAAAACATC TGGGAAGAAG TGGGAAAGTA GGGAATAACT
CTGTTAAATA TGATAAAAGA CAAAGCACCA ACAAAGGCCT AGTTCTAAAC TTGTTATAAT
TTCTCATGGG AGTTGTGGTT TGTCACAAGG TTATGGCGGT CCAAGCAAGT TTACAATATT
TTTTAGAATA ATAACTCCCA GAAATATTTT TAAAATAAGG ACCTTTCTTT AATATGGAAA
AAAAAAAGAT GAAATAGAGA GGAAGAGGTT GCCTTTCCTC AACTGGATTT GCTGGTGTGA
TATGCAGCTG GTGGAAACCA GTTTGTCCAT GCTAGCTGTT GATGGCCACC AGTCCACTGA
ATTAGTGGAA ACCCATGCTT AACCAGTTGT TAACTATTTT TAGTATCATT TCCAGATGCT
ACGCATTTTT TTTGTAAAAA ACATACTTAC AATAGTATCT ATATAAATTT TTAAAATTGT
AATGTATATG TGGTATATTT AACCTGAAAA TAATCTTTGG TGTATACGTT GATGAGCCTA
GGCCTTTGCA GACCTCTTAC AGTTACTCTG TCACAGCCCT TCAAAGGTTC TTTGACTTCA
GATTAAGAAT CAATTGCATG TGGAATGCAT GTGCAAAGGA AGAGATATTC AAGGCAATTG
TTACCATGTA CCATAAACCT TGTACATAAT TTTTTGTCAT TTTCTTTTCC TCTGTCTCTA
TCCCTTCTTC TTTGACACAC AATAGACACC TAGTCAATTT ATTACAAAAA AAATGAATGA
ATGAAGTGGA ATTCAGTTGG GAAATAGGTT AAATAAATTA TTTAGGAGAT GAGGAATAGG
TAAAAAGAGA CAAGTACATA GTTTATTCTT TTGACTTAGA AAACTTTTGA TTCTTAAATT
CTGCAGAATT GGAGAAACTG GTGGGGAAAC TTCTAAAATC ATTATTTAAT TACCAGAGAT
GTAATAGATA TAGACAAAAG CAGTTTTCTT CTTTTATTAT TTTTTCATCA GTTAGTTCTT
AGCTTAAATA GTAGTCCAAA GCTGGTAGGG ACAGAGGGAA TTAGCTGGTG GCTGAATGAG
GAATTGTATC ACTTTTTGTG AATCACGGTG TAAGCACATT TGGTGTTTTG CCATTGCCTA
AGAACATTAG TCACATTAGG TCAATAGAAA ATCACTTTTT AAAGCCAAAT AAAGTTATAT
GTGTTCCCAA CCATGAGTTG GAAAGAATTA ATATATATGC TGTTGGAGGG TAGAACCCTG
CCTAATCATA TGGTTCTGGA TGGCATTGAT CGAATCCTTA TTCTTTCATT AGGAATAACA
ATAGAAAAAA TACTCCTGCC CTACTGATTT CAGGATATGT CTATTTTAAA GTGCCCATTT
GACAAAACCA TTATCAGGGC CATGTTTTCT TTTTCTGCAG AAAAATCAAC CACTCTGGTC
AGTAGTTAGG TCTTATGACA AGCACCATAA TTTCCTTAGG CAGAGTAGAA TATAATAGGA
TACTTCTTTT TGAAACTTAA TATAATCAGG TAGTTCCAGA TAAACATAGC TTGCAAAGTG
ATAAAATACC ATGTTATTTT AGTAAATCCA ATTGCAAGAG TGATGGGAAA CAGAGTTTAA
AAACTTAAGA AAGATATTAA GATGGAGTTG ACTTTGAATA ATAAAGTCAT CCACTGTTGA
TGGGTGACAT TTATTAATAC AGAAAGTTTA CAGATTTTAC CATAAGCATC AGGGTATTTC
CTGCAGCTGG GGAAACCGTG CTTGAAAGGA TGCGTAACTC AAGGAAAACA CAAGCCCATG
TAATAAGTAT TGCATGTGAG AATTGTTCCT ATAGAATTAG AAAGCATCTT TTACATTAAA
ATTTATTTTT GTAAAAAGAG AAAACATCAA AACTTGAGTA GTATTTGCTA TTCAAAGAGT
CTTACATAAA CGAAAACATA CTCAACCTAC TGCCATTACA GAAATATTTG ACAAATTCTT
GCCATGCTTA CCTGCCATCG TTGTTGTTAT CCTACAAATC AATTGGATTT TCACGCCTCT
CCACTGACTG GAACCCTACA ACTTGCTTCC TTTTATCCTC TTTATATATG CTTCAGATAT
GCTTGAAGTA GATTGTTTCT TATTGTTCTT GCTGCTCGAG TTTGTTGAGT AGTTGGTACA
TGCAGAGTAT TTGTATGTTA TGACATATAT AGGTTTTGTC CATGGTTCCT GGCTCATAAC
TCACTCCCAA GACCCTTGTT ACAGAAACCA GAATCTCTCT CTCTGATCTT CTCCTACCCT
CCTTTCATCT GCCCACTGCA GAACTCTAAT CTGATTATGG TTTCTAAGAC CCTCATACCA
GAGAGTATTC TGCCCCATAC CATAGCAGAA GGAACACTGC ACAGAGACAC CAAGAAGAAT
CTGAACAGAC AGGCCTTGTT AGGTTTAGAT CATGTCCTTA TAACCTAATT ATATTTTAAC
ATGGTTATCC ATGCTTTAAT CATGTGTATT CAATGAACGC TCCATAAAAG CCCAAGAAGA
ACAGATTTGA GGGAATTCTG AAGCGCTAAA CATGTAGGGT CTGACAGGAA GGTGAAGAAG
AACTCATCAT GCTGGAAGGG TGGTCCACCC TAACTCCGCA GGGACAGAAG CTCCTGTGTC

TGGGACCCTT CCAGAACTTG CTCTATGTAT CTCTTCAAGT GCCTGTGTAT TTATATCCTT
TCAAATATCC TTTGTAATAA ATCATAAACA TGTTTCCCTG AGTTTTTATG CCACTCTGTC
AAATTAATTG AACCCAAAGA GGGGGTTATG GGAACCCCAA TTTGAAGCCA GTCCATCAGA
AGTTCTGGAG GCCTGGACTT GAGACTGGTG TCTAAAGTGG GCAGGCAGTC TTGGGGACTA
AGCCCTCAAC CTACGGGATC TGAAACTGTC TCCTGGTAGG TAGCATTGGA GTTGTACTGG
AGGGCACTGA GCTGGTGTCT GCTGCAGAAT TGATTGCTTG CTTGCTGGTG GGGAGAACTT
CCTACATATT TTGGGGTCAC CGAAGTCTTT TGTATTGATT GTTGTTGCTG AGCTCCTATT
GCTAGAAACT TGCTTTATTA AGTTCTGTTT ATTCAGGGCT TATCTGAAAG AGAAACATTT
TTTATGATTT GAGATTTCTA AGCCATTTTA AAACTTGGTT TTAACAGCTT GAGAAATTGG
GGGTATAGTA GGGATGGAGA TACATATATT TGGATGTGAC TTCAAGCTAG ATAAAAGTTT
GGAAGGAATA AAAGTTTGAC ATCTAAAGTT TTTGCATAGT TTGAGTTGAG CAGGAGGTAC
TAGGTATGTT TTTAAAATAT TTTTTTCAGC CAGGCACGGT GACTCATACC TGTAATCCCA
GCACTTTGGG AGGCCGAGAC AGGCGATCAC CTGAGGTCAG GCGTTCGAGA CCAGTCTGGC
CAACATGAAG AAACCCCGTC TCTACTAAAA ATACAAAAAT TAGCTGGGTG CGGTGACACA
TGCCTGTAAT CCCAGCTATT CAGGAGGCTG AGGCAGGAGA ATTGCTTGAA CCCAGGAGGC
AGAGGTTGCA GTGAGCCGAG ATCACACCAT TGCACTCCAG CCTGGGTGTC TCAAAATAAT
AATAATAATA AAAAAATGAA AGATTTTTTC TTACTCAGCA TCCTCCAGGC ATTTTATTAT
CTGAGCACTT TATGGGAGTT GCATATTACA TTTAGGGCCC ACTCAGGTGG GTGGGTATCT
AAGCATTTGA AATAACCTTA TGTAAACTAA TAAGGAGTAA TCAGGCCTGT GGCAAGATGG
AAACAGTCTT AGAGGCATTC AAATTCAAAT TTCCTTTAAA ACACTGGGCT GGCCAAAACA
AAAGACAATA CTATCTACAG GCCAGTTTCT AAGATTATCA GATTTTAGTA GCATTTACCA
TTTCATTGTA CTTGGCACAC TTTAGCAAAT TTGCACTTCT TAAAAGTACC TGCAGGCAAT
CTCCTATATA AAAACACAAT GCAGGCTAGC TTGGCTCCTG CCTTTAATTC CAGCACTTTG
AGAAGTTGGA GACTAGCGTG GCCAACGTGG TGAAACCTCA TCTCCACTAA AAATACAAAA
ATTAGCCAGG CATAGCAGCG CACGCCTGTA GTCCCAGTTA CTTGGGAGGC CGAGGCAGTA
GAATCACTTG AACGCTGGAG ACAGAGCATG GAGTGAGCTG AGATTGCACC ACTGCACTCC
ATCCTGGGTG ACAGCGTGAG ACTCTGTCTC CAAACAAAAC AAAACACACA CACACACAAT
GTAACAACAC GAAACAGAAT ACTGTGAAAA TGCTTAATTA TGTCTGACTT TACATGATGG
CTGGATATGT GATTATTTTT TCTTCTTTAT GCTCTTATGT ACTTTGTTCA TTTTTAATGA
TGATCATGTA TAAAGCTCCT CTGTGTAGCA TTCTCTCCAC CAAATTGCCC AGAGACAGGA
AGTCCTGTAA AACAAACTAA GCTCCAAAAA ATGACCTCCT GTTGAATAGG CTTTTTTTTT
TTTTTTTTTT TTTTTTTTTG AAATGGAGTC TAGCTCTGTC TCCCAGGCCC TCGCTCCTTC
CACCTCCTGG GTTTAAGAGA TTGTCCTTCC TCAGCCTCCA GAGTAGATTG GATTACAGGT
GCCCGCCATC ACGCCCAGCT AATTTTTGTA TTTTTAGTAG AGATGGGGTT TCACCATGTT
GACCAGACTG GTCTTGAACT CCTGACCCCA AGTGATCCGC CCGCCTGGAC CTCCCCAAGT
GCTGGGATTA CAGGTGTGAG CCACCACGCC CAGCCTGAAT AGGCTTTCTA ACCTACTGTT
TCTCATTTTA CTTTCTCTGA GGCAGTAAAA GAAACTGACT CTAAAAGGGA GCAGTAGAGA
AGGAACTCAG ATTTTATTTT GAAGATTAAG CTACTCAAGG GCTGAGGAAA TATGTAGAGG
GGAAGTAGAC TCATTATGGC TGGAAATTTT ATTTGGTGAT AGTGAAGCAA ATCTTAGGGC
TTTCTAATTG AGCTCTTGTT TGAAAGGCTC CAATCTTAAT AGAACTATAA GCTAAAAAAA
ATGACCATCA GTGTTTCTAA AGCAAGTTGC TACTCAAAAC AAGAACACTT TGGGAGGCTG
GGGCTGGTGG ATCACCTGAG GTCAGAAGAT GGAGTCCAGC CTCAACATGG TGAAACCCCT
CTATACTAAA AATACAAAAA GTAGCCGGGT GTGGTGGTGC ACGCCTATAG TCTCAGCCAC
CTGGGAGGCT GGGGCAGAAG AATCGCTTGA ACCCGGGAGG TGGAGGTTGC AGTGAGCTGA

CATTGTGCCA CTGCATTTCA GCCTGGGTGA CAGAGTGAGA CTCTGTCTCA AAAACAAAAC
AAACAGAAAC AAGAACTCTA TGTTGAGAAA TCCATACTAG AGGGTTTAAT TTCTCACTTT
GTGTGCTAGT GATTAATAAA TTTCAAGCTT ATCACACAGC CAGAAATGTC GCCTGTGTTT
CTACAATAAA AGATTTGGGA ATATTGTGAC ATTTTTTCAC TGAGCCTTCT GGGTTCATTA
TTAAGATATA ACAATTTTAG AAGACTTATT GAGATAGGTA TACTTTTTAA AATTTGGATT
CAATATATCA AGCTCAGCAT TTTGTCTTTT TTTTGTTTTG TTTTAAACAA CTTTGGGTTT
ACTTGTAAGA GTATTTCAGT GGAGGAAAAG TATCAGAAGT TTAGCACTGA GTGCCTGCCT
AACTATTTTA TGTCTTCTTT CATGATGATG CTTAATCTCA ACATAATGAC TCAGTTTACC
ATTGTAAACT CTTTGTAGGA TGTCACTGAT TTTATCAGTC ATTTGTTCAT GCAGCTGGGA
ACTGTGGGGA TATATGATTT GCCACATCTG AGGAACAAAC TGGGTGAGCT CCATTGAAAA
ATTTCTAACT TTCTTATTAA AAGCAAGCTT CCTCACTTCC TGATGGCTTA TGCAAATTAA
TGCCTATTTA TTCATCAGAG GGTGACTGGA GAGATTATGG TTTTCTTATG AATTTCTTGC
TTGAGCCTGC TTGTCTGCTG TTTCTAAGTT GTCTGCACTA TTTTATGTGA GTAATTTTCT
ACTTTATTAT GTTCTGTTTT CACATGTCAA ATCAGCTCTC CCAAGAATGC TACTTGTAAC
CTAAGTAGAC GTGAACCGAA AAGGGTAAAG ACCCAGCTAA AAAAAAGGTT GACTCAAGTT
CAGTTCACAT TCGTAAAGTA GTTAGGCTGC CTTAGGTTGC TTATTCTTTT CTTGAAAAGA
ATCCCTCAAG AGAGAAACAT GTGAGGCCAC AGCAGCTTAG ATCTGTCTTC CACAGAGAAG
GTGGCTTTTA CAGAGAAATT GACATACTCA TCACTTATCT GACATGACCC AGCTTTGTAA
AACTGGCTTC TATTAAAATA GTCTTAATAG ACTATTCACT GAGGAGGAGA GAATCTTATT
CACTCTTTAC ATTCTCTTCA CATTTTCAGA ATAGATGTTT AAATCATTGC TCACACTGGA
TCCATAAATG TCTAAAATGT TGATGAAGAA ATAGGTGATT AGAGAGTAAA ATTAATAGAG
ACTTACCCTT CCTTGCATTT TAACATAATA TTCTTTCCCC TTTTCCTTCC TCTGTTACTT
GGCTCTTAAA TACCAGAAGT GAGATATGAA AAAGGGAACT GGGAACAAGT ATTGAAAGCA
CCATAGGTTT ATCTTATATT AGCATTTTCC AAACTTTATA ATGAACCAGC AGTGACCCAC
ACATCTGCCA GGTAGGAATT GCTCATCAGT TGTGTCGCGT TATCTTGTTA AGCTTCAGCA
ATGTTCCATA CAGCCACTAA TAACAGATCA AAGTGAGCAT TAGGGTTGAA ATTAGTAAGC
TGTTTCTTCT CAGTTCTTTC TGGTAGTTGA ATAAATATAG AATGTATTAA ATAGTTTTCT
TTATTTCAGA ACTTCTCAGA GTCTGTAATA TATTGTATGG TGGTAGCTTA GGAGGAAGAT
GCAATAGGAA ACTTTTCCCA GATAGGTTCA CTATTTTTTT TTTCCACGAA AAATAAGCTG
TTCTCAAAAT ACAGTTTACA AAATTTTATC CTTAACTCTT CACTCTTTCT CCTAGTTAGG
GAGACCGCTC CACCAGTAGA AAAGATAAAC CCTGGTAATT TGTTGTGTAA ATGGGATAAA
TAGCCTAGTA CCTAGTCATG TGGATTCAGG CAGCACTGAG CCTAAATTAA AGTTTGCAAG
GTATACATGT TAATGTATCT AAGTTACTAT ATTTAGCCTG TTTCTTAAGT ATGTTTCAGA
AACATATTCG TTTTTTTCAG TGGCAGTTAC CTTCAGATGC ATGTGCTTCT AAAGCATGTT
GGTTGCATGT GCAGTACATG TTGCTTAAGC ATTTAGCTTC AGAATGGCAT CTTTCCTGTG
AATGTCTTAA CATTTACAAA AATATACCAG GATCTCAAAT ATCAGTGCTG GTATTTTTTT
TTTTTTTTAC TTAAAGAAAC TGATATGATT AAATATTAAG AGACAATATG ATCCTTGTTG
GCTTGTAACC CTAGTTTTTA TTGTCTTGTA GTTATTAAAT AGAGCATCTG TTGAGGGACT
CTTTTAAAAC CACAGCCATG AACAGACGTT GGGGCTAAGA GACAGAGCAG CCTGCGACAG
TGTGGACCTA CCTGTAGCAG CTAGCAAAGG CCTCTAGCAG CTACAGTCCC TTCTGGAGTC
TTTATTTGCA TGCAAAATGC AAAGGAGTCC TGGTGACCTA CCTCCAAGGC AGCTGCCCTC
CTGAACACTC CCTTGGAAAA CAGTAAACAT CATTTTGGAA TGTGAACAAC CAGAGACTAC
ACAGGAGAAA GGAAAAAAAA ATTCTGAAGA TGCAAAATCT TGGGTGGCTT CACCGTTCAG
TTTTTTAATA AAAGGAACAA TATACAACAC GTTGTTCTTT TTCTCTTTTG AAATCCCTTC

TATTACAGTG ATTTTTTTCT AAGATTGTCA GGATTTGAAG TGTATGTTTT GTTTTATTCA
CAGCGTAAAT TTTATTCACA GTTTAACTGT CTGCCTGAGT GTCTTTCCTT TCTCTAATTA
CCTTGAGGAA CCCAAGAGCC TGTTGTAAGG AGAAAATAAG GCCCTTGGAT CTCTTGAGAT
TCACAGATAT AAGTTATTGA AGGGAAGATG GTCCTATGGA GGACATATTT AAAGAAGGGA
AAAAAGAGGC TTTCTCAGAT ATGTCAGACT GCTATAGTAT ACTTCTACAG ATTATAGACC
TCCAGTACCT CTGGCCAGAA AGATGGTATC GTAAACACCC TATTTTTTTT CTTTTCTTTT
TTCATTAGGT ACAAGCTTTG TGCTAAGAAG TTGACATACT ATAAGCTACA AAAGTTCTGT
AAAGTAGATA TAACTAGTTT CATTTTATAG ATAGAGAAAA TTAATCTCTT ACAGTGCTAA
GCTCACAGAG TTTCTAACTG TAAAATGCTA GAACTTGTCT TTCAAGCCTA AAGACTTCCT
TGGGGCTAAA TAGTGAAAAA AGCCATTTCA CAAATAAGTA AATGGTATTT AGAGGCATAT
TTGGATTTCC TGGTAAATTC CAGTCTGTGA GCATCATGAA TATTAGTTTA ATGTTGCATG
GGCTCATGTT GAAGTTTTAA GAGAAGAACT GCCTTGAAGC TTAGGTTTCC TTAGCTATTA
GGCTACTGAC TTTCTTGCCT AAACCAGGGT TTTTTCATTG AAGACCAAAA CTTACCTTCT
CCTTCAGTTT GTAGTTTGGA AATTGGTAGA AGAGCTTTGT AAACTTCAAA TTAAGTACAA
ACTAAGTGTC ATAGTCAAAT TTACTAATCT TAATTACAGT ATTGTTCAAC TGATTGCTAT
CTTCTAGCTC TTTCCTGCCG AATAATGGTC TTGTTTCCTG CTCTGTTGGT TTAGAGCTGA
CTTCTTTCAG CTTTGGTAAG CCTGAAATTA TGGGGTTATG TTTAATTCAT ATTGTCTGGG
TGGACTTTCC TCTCTTGCAT TTCTGCTTGA ATAGAAGAAT TTTTCTCTAG AGAGTAGTTT
GTCATCCTTA CTCTGTTGAT TCAGATGACT CTTTGTATGA TCTGAGAGGT ATACTGTTCT
GCTATTCTGA GAAGAAGTAT TTCAGAAAGA TGAATTAAGA GTACAGTGGA CTGCTCCCAC
CTGGAAACTT TTATCTATCT CACCTCTGGA CCTGATAAAT TCTTTATCAC TCAGGACCTT
GATGACGCTG CTCTCTGAAA CCCTCCCCAG CTCTCTCTAT TACCGTGAGA AACATCAGAA
CTTTGGTTCC CATTGCATAT CGCAGGTACC TCTGCTTTCA TGCCATGCTG TAATGGAGTG
ATTGGGTAGC ATGTTTTCAT CTCTTTCCAG ATTGAAAATC TGTATTTCTC CCTGTATATC
TTCAACACCT AATGCACATA GAACTTTGTA GGTACCTGGA AAATGCACCA CAGTTTTCTT
TTCTTTTTGC AGACTTTTCA CAAGTATTAC CAACTTACAA AGAATTAATT TTGTAGGATT
CTAGAAAGAC AAATCAGGAA TGGTGCCATA TACATCTTTT TTGATTCCCT GCTCTAAAGA
ATATTATCAG GTTACCTTCC TGCAGAGTTT TAAAAGAATT GCATATTTCA AGCTGACTTT
CAGGATGTAA ATATAACCAA AGCAACTGAT ATGTAAAAAA TATATTCAAT GGCATTCCTA
GATTTTCTTC TAGGGTGTTT TATTGTTTTG GGTTTTACAT TTAAGTCTTT AATCCATCTT
GAGTTAATTT TTGTATAGGT ATAAGAAAGG GGTCCAGTTT TAATTTTCTG CGTATGGCTA
GCCAGTTCTC CCAGCACCAT TTATTAAATA GGGAATCCTT TCCCTATTGT TTGTTTTTGT
ACGGTTTGTC AAAGATTAGA TGGTTGTAGA TGTGTGGTCT TATTTCTGAG ATCTTCATTC
TCTTCCACTG GTCTATGTGT CTGTTTTTGT ACCATGCTTT TTTGGTTACT GTAGCCTTGT
AGTATAGTAT GAAAGATAGC ATGATGCCTC CAGGTTTGTT CTTTTTGCTT AGGATTGTCT
TGGCTATACG AGCTTTTTTT TGGTTCTATA TGAATTTTAA AATAGTTTCT TCTAATTGTG
TGAAGAATGT TAATGGTAGT TTAATGGGAA TAGCATTGAA TCTGTGAATT GCTTTGGGCA
GTATGGCCAT TTTCATGATA TTGATTCTTC CTATCCATGA GCATGTAACG TTTTTCCCTT
CGTTTGTGTC CTCTCTCATT TCCTTGAGTA GTGGTTTGTA GTTCTCCTTG AAGAGATCCT
TCACTTCTTC TGTATTCCTA GATATTTTAT TCTCTCTGTA GCTATTGGGA ATGGGAGTTC
ATTCATGATT TTGCTCTCTG CTTGCCTTTT GTTGGTGTAT AGGGATCCTG GTGACTTCTG
CACATTGATT TTGTATCCTG AGACTTTACC GAAGTTGCTT ATCAGCTTAA GAAGCTTTTG
GGCTGAGATG ATGGGGTTTT CTAGATATAG GATCATGTTA TCTTCAAACA AAGACAATTT
GACTTCCTCT CTTCCTATTT GAGTACGCTT TATTTCTTTC TCTTGCCTGA TTGCCCTGGC

CAGAACTCCC AATACTATAT TGAATAAGAA TGGTGAGAGA GGGCATCCTT GTCTTGTGCC
AGTTTTCACG GGGAATGCTT CCAGCTTTTG CCCATTCAGT ATGATATTAT CTGTGGGTTT
CTCATAAAAA GCTCTTATTA TTTGAGATAC GTTCCTTCAA TACCTAGTTT ATTGAGAGTT
TTTAACATGA AGCGATGTTG AATTGTATCG AAGGCCTTTT CTGTGTCTAT TGAGATAATC
ATGTGGTTTT TGTCTTTAGT TCTGTTTATG TGATGAATGA CGTTTATTGA TTTGCATATG
TTGAACCGGC CTTGCATCCT GGGGATGAAG CCAACTTGAC TGTGGTAGAT AAGCTTTTGG
ATGTGCTGCT GGATTTGGTT TATCAGTATT TCATTGAGAT TTTTTGCGTC GAAGTTCATC
AGGGATATTG GACTGAAGTT TTCTTTTTGT TGTCGTATCT CTGCCAGGTT TTGGTATCAG
GATGATGCTG GCCTCATAAA ATGAGTTAGG GAGGAGTCCC TCCTTTTCAA TTGTTTGGAA
TAGTTTCAGA AGAAAGGGTA TCAGCTCCTC TTTGTACCTC TGGTAGAATT CAACTGTAAA
TCCATCTGGT CCTGGACTTT TTTTCATTAG TAGGCTATTT ATTACTGCCT CACTTTCATA
ACTTGTTATT GATCTATTCA GGGATCCAAC TTCTTCCTGA TTCAGTCTTG GGAGTGTGTA
TGCATCCAGG AATTTATCCA TTTCTTCTAG ATTTTCTAGT TTCTTTGCAT AGAGGTGTTT
GTAGTATTTG CTGTTGGTTG TTTGTACTTC TGTGAGATCA GTGGTGGTAT CCTGTTTATC
ATTTTTTATT GTGTCTGTTT GATTCTTCTC TTATTTTTGA CAAAGCTGAC AAAAAGAAGC
AATAGGGAAA GGACTCTCTA TTCAATTAAT CCTACTGTAT ATCTGGCTAG CCATATGCAG
AAAATTGAAA CTGTTCCTGT TTCTTAATCC ATATACGAAA ATCAACTTAC GATGGATTAA
AGACTTAAAT GTAAAACCCA AAATTATAAA ACCCTGGAAT AGAATATAGG CAATATCATT
CTGGACATAG GAATGGGCAA AGATTTTATG AGAAAGACAC CAAAAGCAAT TACAACAAAA
GCAAAAATTG GCAAATGAGA TCTAATTAAA CTAAAGAGCT CTGCACAGCA AAAGAAACTA
CTGTCAGAGT GAACAGGCAA CCAACAGAAT GGGAGAAAAT TTTTTCAATC TATCCATATG
ACAAAGGTCT AACATCCAGA ATCTACAAGG AACTTAACAA ATTTACAAGA AAAAAGGAGC
CCCATTAAAA AGTTGGCAAA GAACATGAAC AGACACTTCC CAGAAGATAT TCATGTGGCC
AATAAACATG AAGAAAAGCT CAACATCACT GACCATTAGA GACGTGCATA TCAAAATCAC
AATGAGATAC CATCTCATGT CACAATGGTG ATTATTAAAA AGTCAAACAA CATGCTAGTG
AGGTTGTAGA GAAATAAGAA CGCTTTTACA CTGTTGGTGG GAATGTCAAC TAATTCAACC
ACTGTGGAAG ACAGTGTGGT GATTCCTCAA GGATTTAGAA CCAGAAATAT CATTACTGCA
TATAGACCCA AAGGAATAGA AATCATTCTA TTACAAAGAT ACATGCACAT GTATGTTTAT
TACAGCACTA TTCACAATAG CAAAGACATG GAATCAACCC AAATGCTCAT CAGTGATAGA
CTGGAAAAAG AGAATGTGGA ACATAAACAC CATGGAATAC TATGCAGCAA TAAAAAGGAA
TGAGATCCTG TCCTTTTCAG GGACATGGAT GGAGTTGGAA GCTGTTATCC TCAGCAAACT
AATGCAGGAA CAGAAAACCA ACCACCACAT GTTCTCACTT ATAAGTGGGA GCTGAACAAT
AGAACACATG GGCACAGGGA GGGGAATAAC ACACACTGGG GCCAGTCAGG GGGTGGGGGG
TCAAGCTGAG GGAGAGCATT AGAAAAAATA GCTAATGCAT TCTGGGCTTA ACCCATTTAT
GCCTAGTGTT CCATTTCTGG AATGCTAAGC ATGTGGAAGT TCTTTATATC CTGCTCAAGG
TCATTGCCAA GGTCTGATTT TTCACATTCA ACAAATTGCA ACCTCTGGCA TAAATGGGTT
AATACCTAGG TGATGAGTTG ATAGGTGCAG GAAACCACCA TGGCACATGT TTATCTATGT
AAGAAACCTG CACATCCTAC ACATGTACCC TGGAACTTAA AAAATTTAAA ATATATATGT
ATATATATTT AATATGGAAT TTTAAAAATT ACTAATGAGT TCTTTTATCT GAGTAATTTT
GCATCAACAT GCTTTTATTA TGGAAGAGAA GATTCAGTGA GTACAAAATT GCAGATACAT
GTGTCAGAAG ATCCCTGAAT ATAATAAGGC TTAGTATTCT GTGTCATAAT TGCCTGTTTG
TATTCCTCTC TGGTCTTTAA ACTTCATTAG GGCAAGGATC AACTCCATCT TACTAACCAT
TTGATTCCCT ATGTATTACA CGATATATGA CCAATAATAA GCCTTCAATA AATACTTGTA
AAATAAAGAA TGTTATGTAA TATATCATGT GGTATTGTTT TATTGATGTG TTCTTTGAAG

TGTTACCTTT GTGTCCTTAG AGGCTATGGT GGGCCCTCAG GCCATATATG TTGGATGATT
TAGTGGAAAA GAGATTAGAC TTTTATATTT TATTCTATTT ATTTATTTAT GTTTTAATCT
TGTGTAAGCC TATAGAGGCC AAGATAGGGC TCAGAGTTAA AAACAGTGAT GTTCTAACTT
TCAAAACTAC TAAAAACAAA ATTAGCTATC AAGGGAAATC CTGAGTGTTC TGACTGCAGG
GGTTCCCCAA AGCCTGGAGG ACCGAAGCTT TGAAATGCTG CATGGTGTGC CTCACCTCTG
TTGAGTATAG TGAGGAACAT GAGCTGAGGA GTGGATTAGA TGAACTCTTA AGATTTTTTT
CCAACTTTAC GGTTTGAGTT TATGAGTAAG TAAGTGAGGA ATGAGCCCAG GGACTTTGAC
TACTTTGCTG TGCAGTAAAT TGCTGTTTTC TTAGCTAATT CTGGAGTGTG TATTTTGGTT
TTAATGTAAT GTCCTTTCAG CACTGGAAAT AAACAGATTT ATGACCCTCC ATTTTCACAT
GTAACTTCCA GTTCCTTCAG TAAAATGCTG GAAGTTGCTT GGCTGTAACC ACTGTAACTA
TTTCCTGGCA GTTTTGCATT TTATTACAAT TCGACTTTTT AAATGCATGG TTCAGTCCAG
TTAATAGGCA GAAAACTATA GCTATTCCTT GGAACTGACA TAGGGAAGTT GGTTTCTTTC
TGAGGAAAGA TGTGACACTC TGATGGATTT GGCAAAATAC CAATTCAGTA AAATTTGCCT
GGTTCCTTTC TGACAGCAAA TGGAGGTAAT TGTACGTATA TAGATTTCTA CAGTAAACAA
AGGGCAAAAA TGGAAACATT TTTTGTTTTT TAATGTTGCT GGAATATAAT ATTTCCATAT
TTGGTTAATC ATCAAATCAT TCAACTCCAT GCAAAGTACT CTTTTCAAAA TTACAAATCA
TCTTGATGGA TTAGACTTGG ATGTCTAAAT TTTTTATTAT GTATTCTAAA TTATCTATAG
GTACATTTTG AAAATTATAT ATACATGTAT CACTGTACTA CATTTGTTAT AAAATGTGAA
ATATTAAGCG ATTAAATACT TTATTAATGT CATAAAATTG TGCCCTACCC AAAATATATT
TTAAGAAGAT ATCATTTGAA AATAATGTAA AATATTCTTC ATTATTATGA TTTTGTTGCC
AAAAGTAATG TTAGAGTCCC TGTTACTGAT AATTTTCATG ACTGTAAATT TTCAACATTT
GGAAACAAAA TTTTCTGATG TGTTTTAAAA TCTTAAAGAA TTTCAGTAGA TTTTGACACA
TTTTTGTTAA AAGTATGAAA GTGTCACAGC TTTAGAAGGA TTTAGAATAA TTATCAATTG
GTATTAGGGA TGACAATTTT AATAGACAAA TTTTAACTGA AACATTTGCT TTATTGATGT
CTAGGACTAA AAAAAATATT ATAGTGATGG AACCATTTCA AAACATCTGT TTTTAAAAAC
TTCTGTCAAC ATGTTTGACA GAAGTCAGAA TAGGGTAATT ATAAAAAAAC TGCCATTTTC
TTATGTTTGC ATCATTAGTA TTTGCAAATC CTTGTTTCTT TGCAAGTATA TTTTAAACTA
CTTGTTCATT TTATGAGTGT TAGCCTGATT AATTCAGATA AGATGCTCAA GTCAGTTTAA
GCAGGCAGAT GTGAAAAAGG GAATATGCTG TGCCTTAACT ATCTTATAGG GACTTACTAG
CACACTGTAC ATAACAGATG AACATCTGAT ATTAGATCTG ATGGGCACAC CAATGTTGTT
TCATATATTA AATAATAGTG GAAGTTTTCA ATTTAATTAG TTTATTTATT TATTTATTTA
TTTTTGAGCC GGTGTTTCCC TCTTTTTGCC AAGGCTGGGG TGCAATGGCA CATGGTTGCA
GCTCACTGCA ATCTCCACCT CCCAGGTTCA AAGGATTCTT CTGCCTCATC CTCCCAAGTA
GCTGGGATTA CAGGCGTGTG CCACTATGCC AGGCTAATTT TTGTATTATT AGTAGAGACA
AGGTTTCACC ATGTTGGTGA GGCTGCTCTC AAAATCCCGA CTTCAGGTGA TCCACCCGCC
TCAGCCTCCT AAAGTGTTGG GATTACAGGC ATGAGCCACT GCGCCTGGCC AATTGTATTT
TTATGTTAGA AATATATATG GGAGTTACAA TACTTTGTTT CCAATGTTCC GCTGGAAACC
ACTGGTTGAA AGGATAGGTC CCAGTGTCAG CTGCCCAGTT GGCAGTTTTG GATAGAATAT
AAGTTCTATG AAGGCATGGA TGAAGTATCC TTCATGGTAA TGATGGTAAT GTTAGTATAG
TCACTTGTTT AATGTTGTAT AACTTTTTAA CTCTTCCAAC TTCAATTTCA TCTATATCTT
TCATAGCAGT CCAGTAAAGT AGTATCAATT ATAGATATCA TTCCTACTTT CAGATAAAGT
AACTGAGGCT CACTAGTTTT GACTTGCTCA AGGTTACGTA AATGATAAGT GACAGACTGG
ATTTAAAGGG CGGAGTCCAC TTTACTAAAC TGCCTCTGTC TTCTAGCCAA GTTTGTTATG
TAAAGTAAGT ATTCAGACAT TGGCTAAGTA TATTATTAAT CAAGAAAACA GTAGAACAGG

TTTAAGGAAT TTTTTTTTCC CTCTCCTCAG TGTCAAAAAG CAGTCAAGTA AAAGATCCAA
GATAAATAAC TTCAGTTCCC AAAAATAACT TCAATTCCAA AATTTCCCTA TGACTAGGTA
TATCTAATTG GATAGCTGAT AAAATGAGTG GGAAATTGAA GAAGGGATAT TTTGAGAATG
TGTGGCAACT TAATGGTAAT ATTGAAGAAG CAAATGGATA GTTATTGAAT CATAATAGTT
TTTTAAACTT AAATACTAGG GCTTGATCTT CATATTTTAG GAAATTATCT ATTTGTTGCA
GTAGAAATTA GCATAAATGA ATGGAAACTT AAATTATCTC ATCTGTGCAA GGCAGTCTTC
AGAGACTGAT GTTTATATAA GTGAAAACTC AGAAAAATTT TTTGGCTGCA TTTCAAGCTG
GATAGCGGGG TCAAAACTTC AACTTCTTAA TTGGTGATAC AACGTCAGAA ATCACAAAAG
AAGTATTTCA CAGGAAAAAA TGACATACCC CGAAAACATT CAAATCAAGA AAATCACAGG
GGTACAGTGG ATGGTCCCCC TTAAGGTACT GCGTACTTGG CTTCCCCGCT GGTAAATGCC
TCTCTTTTAC TAAGTTGTTG TCGTGCAGAG TGGCATTGCA GCTCTGTACT TTTACAGAAT
AATTTCAGGT TCTAAAATTG ATACATTATT TCAGATTGTA AGAATAGTAT AAAATAAAAT
TTTAAAAGGG AAAATAATTC CATCAATTTA ATGGAATGTG TAGGGTTTAA GTTATAACAA
CAAAAACAAA GTTGTAGTTT TTTAGGATTA CTCATATAAA TAGTGTGTCT ATTAAGAATT
ACTGGCTTTA TGAATATTAA ATAAGAAGGC TGGCCGTGGT GGATCATGCC TGTAATCCCA
GCACTTTGGG AGGCCGAGGC AGACGGATCA CCTGAGGTCA GGAGTTCAAG ACCAGCCTGG
CCAACATGGC GAAACCCTGT CTCTACTAAA AATACAAAAA TTAGCCGGGC GTGGTGACGC
ATGCCTGTAA TCCCAGCTAC TCAGGCGGCT GAGGCAGGAG AATTGCTTGA ACCCGGAAGA
CGGAGGTTGC AGTGAGCTGA GATCGCGCCA CTGCACTCCA GCCTGGGTGA CAGAACAAGA
CTCCATCTTG AAAAAAAAAA AAAATTAAGA CAACCACAGA TATAAGGATT TGCAATTAGA
AAGCTTGAAG GGAAATTTTT GGAACTTCTA GGTCAACCTA AAATTACAAA TAGGGAAGCA
GCTTTGGAGA GGAACATTTG CACTTCTGAA GTAATACTGT GATTTGGACT GTCAGACATT
GTGATATATT GGCATCACCA CACCTCTGTT AATTATAAAC TTACCTATTC TCAGTGCCTT
GTTAACTGTT GAAATAGGAC TTGACATTTG TTGTTCTTCC TCTTCTGTTA TTCATTTATT
CACTCATTTA GTAGTTAATT ACTTCCTAGT GCCAGATAGT TTTCTAGGTA CTGAATAAAC
AAAAATCCTA TTGTCATAGT TATGTTTTTG TGGAAAGAAC AACCTGATAG ACTATCGTGA
AAAATAAATG AGTTAGTATA CATAAAACAC AGCCCATGAG ATATAGTCTG TAATTATCAT
CCCTGCTTCA TTTATTTATT TATTTATTTA TTTATTTAGA GAATGGATCT GATTCTGTCA
TCCAGGCCGG AGTTCAGTGG CTGGATCATT GCTCACTGTA ACCTCAAAAC ACCTGGCCTC
AAGTGATTCC CCCAACCTCG GCGTTCCAAA GTGCTGGGAT TAAAGGCATG AGCCACTGTA
CCTGGCCATT GCTCCTGTTT TAAAGATGAG GAAACTGAAA GTACAGAGAG GAACGTGACT
TGCTCAGGAT CACACAGCCA ATCAGTGGCA GAGCAGTCTA GGCAGTCTAG GCCTGAAGGC
ATTATTCTTT CTTTCTGTTC TGCGTCAAAA ACCCTAGGCA GGCATGGTAA AAAGACTGAA
AGAGCCAAGG CAACATAGGT ACCTGAAGAA TGGGAGAACA CCTATGAGGA CTGCTAGAAA
TTTTAGGCAG CCCTTTAAAG GCCTTGGGCA ACAATTTGAA CGTGACTCAT GAGTCAGTTA
ATTCTCTGCA CCTTTTTTTT TTTAGAGAAT ACAATGGATA AAGCAATGAC CTTTGTAACA
TGTAAACCTG TATTTGAATC CTTTCCTCTC CATTTAATTT TGCAGAGTCA CCTTTTCTTC
ATTAGAAAGA AATTTTATCC CAAAGAGATT ATCGTGAGGC TCAAATAATA TAGGTGGAAG
CCCTTTGGGT TTAAAAAAAA GCCTTTTCTG TCATTTCTGG GTATCTCCCG TCAGTGATTT
TAGTAATTGT TTACCTCTCT GGCTCTTCGT GAGAAGACAA ATTGTAGGAA GAAGTGGAGG
ATGGCATTGG GTAGGATGCT TAGTTCTAGT TCCACCACTT CCTTAGGTCT TTTTTTTGTC
CCCCAAGACT ATATCATTTC TGTGTTCTCT TGAACCTGTA AGATTTATGA GTAATCTGTA
CAATATTTAT AAAGTATGAA ATCTCTGCCA TTGTGGCACA AAAGTAGCCA TAGACAATGT
ATAAATGAAT GGGCATGCTG TTCTCCAGTA AAACTATATT TATAGAAATA GGCAGTCTTC

TCCAATTTTA GTTTGCTCAC CCCTTTCCTA TAGAACTCAG GTTTATCATA TACAATGTTA
CACATTATTC CTTTGAATAA ATTGTTTACC TAATACTTTT TAAATTTGGG CATTTTCTAG
GACATTCTAT TGTTTTTATT TTTAACTTGT GTTATACTGC AGTGAATGTC TTCCTGTGTG
TAGCATTTTG TTTGTGTATG TGTGTCAATA TCTATAAAGT AGCTCTTAAG AACTCTGGAA
ACATTGTCAG ACTGCTTCCT CAAACAGCAG AGGTGGGAGG GAGCATCGGC CTAACCCTAC
CTGGCCTGCC TTACCAGCAT GCAACCAAGG AGAAGAGAAT AGGATCAAGA GCCAGTGCCA
GAGAGAAGTA CATAATACAA CTTTATGGAG TTTAGGTTGC ATCCTATAAA GTTACCATGT
TACACTGTTT CGATCTACCA AAGTAGAAGT TTTACATAGT TTAGTATGAA GTTAATTTTA
TTAACCTTTT CACACAACAC ACCCTCTCAT ATTGCTGCCC AACAATGTCC TGGGTCTTTG
TTTCCATGGT CATCATATTG AGTAGGGTTT TAGCAGGGAC TTGAGTAGTT GGGAACTGAA
CTCATGCTAG TGAAGCCATT CTTCACTAGA AGCAGACATA TGGGATCCAG AATATCATGA
ACAACCAGCT GGGTACAGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAAGGA
AGCAGATCAC CTCAGGTCAG GAGTTTGAGA CCAGCCTGGC CAACATGGTG AAACCCTATC
TCTACTAAAA ATACAAAAAT TAGCCAGGCC TGGTAGCACA TGCCTGTAGT CGCAGCTACT
TGGGAGGCTG AGGCAGGAGA ATCGCTTGAA CCTGGGAGGC GGAGGTTGCA GTGAGCCGAC
ATCTCGCCAG TGTACTCCAG CCTGGGCAAC AGAGTGAGAC TCTGTCTGAA AAAAAGAGAA
TATCATGAAT ATCACCAAAG ATAAAAGCAA GATGCGGATA TGAGTAGCCC AGAGTAACTT
TTCCTGTGCT ATATAACACA GTGAGTGGTG ATTCCCACAA TTCAAGTCCA GTTTTTCTGG
ATTCTAAAGC CATACATTTT TAACTCTCCC CCTAAATATG ATTTGAATTC AGAGAGTCAG
ATGCAATAAG GTAATTTTAA CTTTAAGTGT AGTTTAGCCC ACACATTTCT CTAAACTGGG
GGTTGGCAAA CTACAAACCT CTTTTTGTAA GTAAAGTTTT ATTGGAACAC AGCCTAGCTC
ATTCATTTAT GTATTCCCTT TGCACCTTTT AGGTTGAGTA GTTGTGACTG AGGTCTTCCA
GCCTTCAAGC CTAAAATATT TACTATCTAG TCCTTCAAAA AGAAGTTTGC CAACCTCTGC
AAACAAGTGA AATAATGTGT ATTAGAGTAG AGCAGGAGTC CACAAACTAC AACCTCTGGA
GCCAAAACCA CCCTACTGTC TGTTTTTGTA AACAGAAATT TTATAGAAAC ATAGTTGACC
CCTGTGCCAA TTTGTTTGAG TCTTGCTTTT GATGGCTTTT GTGCTACAAG GGCAGAGTTG
AACAGTTGCA AGAGAACTAA GTGGCCTGCA AAGCCTAGTG GTCTTTGCTG TCTGGCCATT
ATAGGAAGTG TTTACCAAAC CCTGCATTTG ATCATGGAAT CCCAGAATGT TTATACCTTT
GAGGATTTTT TTTTCTACTT TCCATATTTT ACAGGCAACA GAACTAAAGC CTCGTAAAAT
CTTGCCCAAG GTTCTACAAA TAGTCTTTTC TATTAAATTA AGTCAGAGCC TGAAATTGTC
TGTGCCCAAG TCTCTCCTCA GAAATTTAGC ACAGCATTTT CCCTTTCTGT CTCCCACATG
CTACACAGTC ATCCTGGATC ATGTGCCCCG CCTCCATGAG GAGATGTTCT AACATGAGGG
TGACTAGGCT CTGGAGCCAG AGCATCTTAG TTCAAATCCT GGATCTGTTA AAAATAATAG
TTATTTTTCA GTTCTAACAC TGAAATAAGA CAAAAGGCCA TATTACTGAA ATAAAGTGTA
AAGAATATAT TTAATACTCA GGTTACTTTA GCCCCCTCAT GATCAGAATA TGCTAATAAA
TTGAGAGATT GACAGAAAAA AAGGATTCTG TTTCACTGTG GCTCACACCT GTAATCCCAT
CACTTTGGGA GGCCGAGGTG GGTGGATCAC TTGTCTGGAG TTCAAGACCA GCCTGGCCAA
CATGGCAAAA CCCCGTCTCT ACTAAACAAT ACAAAAATTA GCCAGGCGTG GTGGTAAGCA
CCTTTAATCC CAGCTACTCG GGAGGTTGAG GCAAGAGAAT CGCTTGAACC TGGGAGGTGC
AAGTTGCAGT GAGCCAAGAT CACACCACTG CACTCCAGCC TGGGCAACAC AGCAAGACTC
CATCCCTGGA AAAAAAAAGT TCATCATTTG TTCTTGTAGT TTCCTAGTGT GGCCCTTGTC
TCAGTGGAAA ATTGTAGTGT GATAGACTGA GGAGAATTCT GTTTCTGCCA CTTCGCCTGT
TGCCTTGGTA GGAGCTCCTT ATCTCTGAGC TGAGCTCCTT TTGTCCTTGG TAAAATGTGG
TTATTATCAT CTGCATCCTC ATAGTCTTCT TGTAAGGACT GTGAAGTCAC TTTTGAAAAA

TGCCTTGAAT ATGCCAGATG GTCTGTTTTC TTTCTTTTTT GTTTTGTTTT GTTTTGTTTT
CCCTTGGCCT TCATTCCTGG AAGTGTTTTG TTATATACAT TTTGTTTGGT GGCCAACAGT
GCTAACCACA GAAAAGTTTG TATGTCTTCT TTTTCTAGAA TAATTCTTAC AGTGTAACCT
CTCAGGGTGA AGTTTTTGTT TTGAAAGGGG CACATTTACT GTGAATGAGA GCCATGGAAA
TGGACCTGAA ATTTGAAGAC AGATTTTACT TGGTTTTCTT TTCTCTTATT TGTTTGCAGT
ATAAAGAAGA GGGCAGCAAG GTGACCACTT ACTGCAACGA GACAATGACT GGGTGGGTGC
ATGATGTGTT GGGCCGGAAC TGGGCTTGTT TCACCGGAAA GAAGGTGGGA ACTGCCTCTG
AGAATGTGTA TGTCAACACA GCACACCTTA AGAATTCTCA GGAAAAGTGA GTTGCTACAA
ATGAGACATA CGTTTAGTTT TGTTTTATAG TATATATGAA TATGTGTGTA CATTTTTGGA
ATTTTAGTTT GATTATACAA AATATCTTTG GCTTAGAAAT ATTAGGCATG CTATGTAAAA
CCTTACTGGA AAAATAAATT GACCAACATT ATTGAGAGTA TTTTTTCAAA GTGTTCCAAA
AGTAATGGAC CAATGATTAC TTTAAATGAA ATCATGTAAT GGACCACAGA ATTGCAAATT
ACTAATAAAG AAAAGCCATT TTGCTTATTG CCATGTAATA ACATGTTGCA TGATACAAGT
AGATACGTAT GTTTATGCTG CAACAAGTAT AGGTGATACT AATTGGGCAA CTTTTAAACA
AGACCATAAA TAACTGAAAT CAAAGTTCTT AGTATTTATG CAGCCTGTTG GTTTGCGAGG
GCTGCCATAA CAAAGTACCA CAGACTGGGT GGCTTAAGGA AGGTTGTTTT TTCACAGTTT
TGGAGGCTAG AAGTCCAAGA TTGAGGTGTT GTCAGGTTTG GTTTCTTCCA AGGCCACACT
CCTTGGCTTG CAGATGTCTG CTTGCTTACT GGGCTCCTAA ACAGTCTTTT CTGATTGTCC
TAATGTCCTC TTCTTATAAG GGCATCAGTC ATATTGGATT GGGGCCCACC CATATGACCT
AATTTTACCT TAATGACCTC TTTAAAGCTC TGTCTCCAAA TACAGTCATA TGGCTGTGAT
ACAGGGGGTT AGAATTTCAA CGTATGAATT TTGAGGGGGA CACAATTCAG ACTATAAACT
GCAGTTAATG TTTACTGTTA AATTAAATCT ATAACTGACT AAACCACACA AACAGGAGAT
TTCAAATTAG ACTTTATTAG TTTTGGAAGG AAGAGATGAA ATGTTGTTAC TTTTTCTGTT
TTATGTAGTT GTAAAAGCCA CTAAAAATGT ACATGTACAA ATCATCCCAA GCCAGGCAAC
ATTAGTAATG AATGACTGCA GCACAGAGGT AAGAGAGATG ATTTAAAGGA GGATGAATTG
TCTGCAAAGG TGCTGGGCAA GATCTAGACT AATTCATCCC CTTCATTTTA AGTTTAGGTT
TTAAATAGCT TTGTTTGGCT TGATCCTAGA GCTACACATT TACTTTTAAC TTGTTTACTT
TGTACCCTAT CATTTAGGAT ATGCTATGTA CTATTGTACT CTATTTGATA TTTCAAACAT
TCTCTCATTT AGTGAAGAAG CTCCCAACCG GAGTGCTCAG AACCTACATT GCCTCACCTG
TGAAGTGAGC ATGTGGGACT GAATTGCCTT TGAGGTCTCT TTCAGCTCTT AATGATCTGT
TATCCCATAG TAAGATACAT TATTTTTAAT CTCGTTGGAT CTTAAACACC AAAAATAATA
GTATTTAAGA CGTAGGATGC TATCTTGTCA TTATTTTAAT GCACATGTCA ACACACGAGG
TTTTGCTAGA TGTTATGATA AAGCAAGCGA AACAGCAGGC TACTGCTCCT CAAATATCTA
CAGTCAATGA AATATTGACT GGCGTTGTGG GAAAATATCT CAAAAGTATT TTATTTACAA
TTTAAAATAT ATTCTTTTGC CTAAAGAAAC AATGAAACAA CTATAACTTG TTTTTGTTAT
TGTGTTAGTG TAGTCATACA TATGATACAG TTTTATAATC TGCTTGATAA GGGTACGTCT
CTGGGTTTTT CACACTTTGC AACAAGTGGG AAAGAACTTC GACAAAAACA TTAAAACATT
ATCAACCTGG TTGCACGTTT TTCTGATTAT TTTCAGTGCT CTTAGGTTTT TAGACCATAT
CAAAATTCTC TCTCTACACA CATGATTTAT AAGGGAAAGA AAAGTTTGCA GTATGCCTAA
AATGTTCCTC AGGTTAATTA TATCTTGTTT TGCATAGTTG AATGTTTATG GAAGTCTTAG
ATAGTGAGTC ACTCTGAAAC CACAGACTCT TTTTTAATAG TAATAATATG GTGAGTTCTA
TAATTCCTAG CATCTAGCAC ACAGGTATGA TAAATGTTTT AATTAAGTGA ATAATCAGCT
TCCTGAATTT TTCTACTTTG TAAATATGCT CTTTGTGTGT TGAATCTATT TACATATATA
TCCAAGCATT TGGCACAAGG TAAACCAATG TGGCTCAGAT TTTAATGTTA TTGTGAATCA

CTTGGGGATC TTGTTTAAAA AAAAAAAAAA AAAGGAGATT CTGATTCAGG TGATCTGGGC
TGAGTACTTA CAGTATGCAT TTCTAATAAA CTTCTGGGTG TTGCTACTGA TATTGGTTCA
AGAACTAAAT TCTGAGTAAC AAGTATAAAC CGAATGTCCA TAAAGTATAA TTTCTTTTAG
TATCATATGG TTATAATGCT TGCATTTTGT ATATATGTTG CATACAGGTT TTGCTTTTCT
CTTTTTTCTT ATTAAAAAGG TTATAAAGGT ACACTGTAGA AACTGTAGGG AGTAGGGGAA
AGCTTGGCAT CTCATGCCAC CATTCAGAGG TAACCATGTT TGGCATTTTA AAAATGCATT
TATAGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGC GCGCATACAT ATATATGTAT
GTATTATACA TATACGTTAA TATACATATA ATATACATAT ATTACTGTGC AAAAGTTGAT
AAATTGATTA TGATTATGTA TAAACTTTGA TATCTTCTCA TATTTAATAT AATATAGTTG
TGAGCACTCT TCCATATCAT TAAATAACTT TCTAGACATT TTAAATGAAT ATATAGAAAA
TGAAACATCA TTAATCAGTC CCTTATCTTA GAATATTTAT TTCGGTTATT TACAATCTAT
TAGTAGTTTA AAAGTACTTC ATGGAAAATC CTTATACATG CATCTTAGTT AACTTGAACT
TTTATTTTTA GAGAAGACAT GCCAGGCTGT TCTAGGCTAT TGATACATAT TGCCAATTAC
CTTCCTGAAA GATTGTACCA CTTTCCACTT AGGCACAGAG TGTGAATGCC TGTTTTAATT
ACTCTATTTA CATCAGTAAA GACTATATCT GAAATTATCT CATTTTTGGG GACAGGTATT
CTAATAGGCT CTACAAGTAT GATCACAACT TTGTGAAAGC TATCAATGCC ATTCAGAAGT
CTTGGACTGC AACTACATAC ATGGAATATG AGACTCTTAC CCTGGGAGAT ATGATTAGGA
GAAGTGGTGG CCACAGTCGA AAAATCCCAA GGTAATCAAG CACACATTTT ATCATTAATA
AAAATATGAA TGCTGAATAC CATCCTCCCT CCTAAGCAGT CCTACCAGTG TTGCTCGCCA
TTTTATTGGT CATCCCAGTT GTAACTTTTT ATGTGATCTG TTAACGATCT TTATCTCCTA
TTGATACATA TTCTGTCACT CTCCAAGTCC TATTTATGAG TTTTTCTTTC ACATGTCTGT
TGTTTCCATT GCCCCCTTTC CATTTTCTAC ACTCTGGCAT GCAGCTTTGT CCTGTTTCCC
CCATGCTATG GATTAGCCTC CTGCCGGTAA CCACCTGAGG GGTTCTTCCT GCCTGCTGCA
TAAAGAAAAA CCATGGCAAT ATAGTAGAGA AAGAGCTTAA GAGACACTAG GCTGGCCATG
CCACGTGGGG GATGGAGTTC ATATTCAAAT TATCTTGTCC AAAGCTCATA GGTAGGGGTT
TTTCAAAGGC AGCTTGGGGG AAGGGGTGGG GGTGGCCAGG TAACAGATGC TTGCTGCTGA
TTGGTTGGGG TGGGTGAAAT CACAGGGAGT TGAAGCTGTC CTCCTGTGGG CTGAATTGCT
TCTAGGTGGG GCCATAGGAG TGGGGTTGCT GGGTCCAGGT AGAACCACGG GTGTCAGACA
TGCAAAAATA AAATAAAATA AAATAAGATA ANGGTAAGAT AAAATAAAAT AAAATAAAAT
AATAAATAAA ATAAATATAA AATTTCCCTG AAAAGATATT TCAAAAAGCC AGTCTTAGAT
TCTACAATAA TGATGTTATT TGCTGGAGTA ATTGATGGAG TTGCATGTCT TATAACCTCT
GGAATAACGG CTGACAATCT CTCAAGTCTG CGCCTTAGCT GGACTCAGGT TCCTCTTCTC
CCCACAGCCT GACTGCCTCC ATTAGCTTCA CAAAAGTGGT TGGGTTTCAG GGCAAGGCCC
ATTGTCATTT AAACTGTAGC CGAAATGACT TCCAAAGTTA GCTTGGCCCA ATAGCCCAGG
AATATTTAAG TGGAAGGCAA GATGGGGGAT GGGTTAGCTT AGCTCTCTTT CACTCTCATA
GTTTTCTCAC TGGTATAATT TTTGCAAAGG CGGTTTCATG CCTGCCATCT CTTTCGTCGC
TACCTCTCCC AGTTCCCATT CTTAGCTGTT TTATGAAATG CTTCTAGTTT CATCCTCTTA
TACCAAGTTC TGGGAGACTG ATTTGAGTAA TAATAAAACT CCAGTTTCCC ATACAGCCGG
CTCTGCGTGA ATTAAACTCA TTTTCTATTG CAATTTCCCT GTCTTGATAA TCAGTTCTGT
GTAGGCCGTG AGGAAGGAGA ACCCGTTGGG TGATTACGAG ACTGTGTTAC TGCCCACTAC
CTAATGGATA CATTTAGCCT GGTATCCAAA CCCATCTAAT TATGACCATA ACTATATTTA
TCACCTTGCT CTGCTTTCAA ACGATATGAC ACACAATGAA TGAAAACTTT CATTTTTCAT
CTTCATTTGT GCTGTTCCCT TTGCCTCAAA TAGCCCTCAA CTTGCCTACA GTAACTGTAA
AATTTGCCAC CTAAAAAAAA ATCTCAAAAT CCTCTCTATG CTTTGATGTC CAGCAAAAAA

AAAAAAAAAA AAAAAAAAGT TATCAGTATA GCTCTATTAC TCTCAACCAG AGGTGGCATT
TTTCATCTTT ACCCATAAGC CCCATGGTAT ATTTCTTAGT TTGGGAAATT ATATCATAAT
ACTTTGAATC TGTCTAGTCA GAAATTTAGA TAATTTAAAA AAAAATTATT TTTCTGAAAG
TATAGCTGGA TCTAGCTTAG GATCTACATC TATCTAATAT AGTTCCTTAA CATNTTGTTA
AATGGCCACN GGNATAAGTC CTGTAATGCC ATACTTTGCT TTAGGATCAT GTGACTAAGG
GGTAAGGAAT TGGAAAGCAA TGGGGAGCTA GCAGAATTTG AAAAATTATA TAGGTAGGTT
ATTTTTCCTT AATACATTGA AATAGCCTCA AATTCTCAGA GAATACAATG TTTAATCCTC
TAATATCTGT AGAGTTCATG GCTCAAATTT GTATTATTTG AATAAACTAC TAATAGATTA
ATAATTACTC AACTAAAACA CTTTGAAATG TGAGAGTTCC CTTATCTCCC TCGCAGGCAT
ATGACAGAGG TGTGGCTTCT CACACCTTTG GTTGCCCTTG CCCTACCACC CAAACCCCTA
GGGGGAGCAT GCAGAGGGGC AGGTGCAGAG GCCATGGGGG GTGCTTTTGG GCTCTGGCCC
CACAGCAGTG TCTAGGAGTG GATGTTGGAG ACTCCTGAAG CCCAAGTGGG CATGTGTTAC
AGTGTGCTCT TTCAGCTTAG CCGTCTGCAG ATGGCTTGTG TTAATCAGGT CATTAGACCC
CATGCCTTAT TGCAAGGGCA GGGGTCCAAT GTGACAGCCT AAGTTCTTGC TCAGTGTACC
AGAAGAATTG GATCACACGT GGGCTGGAAG GATGAGCACA AGGTTTTATT GAGTGGTGGA
GGTGGCTCTC CGCGAGACGA CTCTCAGCCA GAGAAGGAGG ATGGAGGGGA AAGGTGTTCT
TCCCCTGGAG TCCACTTGTT CCTGTTCTCC TCTGGGTTCA AATACCTCTT CCCTCCTTTT
CTGCTGTGCT GTCCCACCAC TCTCCACAGC TCTGTGCCGC TCTGTTCCTC TGCTCCTCTG
GATGTTCAGC TCCTTGTATC TGTGCCTGCT AAGATTTTGG GTTTATATGG GGGAAGGATG
GGGGGCATGG CGGGCCAAAA GGCACCTTTT TTGGTGTGAA AACAGAAATG CCTGTCTTCT
CTTAGGGCCT TAGGTCTTCA GGCTTGAGGG TGGGGCCTTT GCTGAGGGAC CACCCTCTTC
TACCCAGTAT TTCCCTGTCT CCTGTCCATA TCAACATTAC TAACTTTTTC ATCTGCAGAC
TAATAATGCT AAGGTGTGGC ATTTTTCAAC TGTGAGACTA TGTGAAGGTT TTTCCTGTCC
AACTGATGGC ATCCTCCCAT AATTCTACCC CTTTCTTAAA AGAATCTTTT GCAGTATTTC
TCCAAGTTTA TTCTAGAGAA TTTCTTGTCT GTGAAATGCT CTAGTTAATC AAAATTGGAA
AACGGAGCAT ATCATATCCC CTTCTCAAAT TCACCAAAGT GAAGTCCTAA TGTGTCTTAA
TGTATCTGCA TGAGACAGGA AGCTGAGATC TATTCAACAA CAAAAATCCA AACAAGCATC
AAGAGGAGGA GTGTTAGCAC TTGAGCCTAG GGAGACTGTG GCTCCTGCCT GAAAGATGGG
AGCCTCAGTC ACAGCTGCTT TACCAAGTGT CATATGCTAT GTTTCTGAGG ACTCCTGCTA
AAGCTCCCTT CTCCCTCCAG CCAACCACTT TTGTTTTAGA CAAGGGCTGG GTTTATGAAG
GACTGTTTTC ATGACTAAAG CTTTATAGAA GGTTTAAGAT AAGGAGATGG AATTGAGTGA
AGTAGGAAAT ATGAAAGCAG ATATTATAAT CTGGCTTCCT GATTTTTCAC TAGCATTTTT
GTTTATAAAT TAGTTCTGTT CTAAGAATCC AATGACGTAA TAGAAACTCT CAAAGATTCT
TAACTTGAGA TATAGGGAGT CTTTGAAACT GCTGAAATTA CAGACAGCAT TTATTGTTTA
TGAGCATTTC TGAATCTAGA GCTTTCACTA GATTTGTAAA GAATGTGGGC CAAAAGATTA
AGAGCCAATC CTGTATCTTG TACTCAAAAT GTTTGTAATT CTCACCTTTT TATCCCCCAG
ACTCCTCTTG CTCTCTCTTC ATTTTCACAG TGTTACTGGA AAGGGGTCCT GATCCAGACC
CCAAGAGAGG GCTCTTGGAT CTTGTGCAAG AAAGAATTTA GGGCGAGTCC TCAATGCAAA
GTGAAAGCAA GTTTATTAAG AAAGTAAAGG AATAAAAGAA TGGCTACTCC ATAGACAGAG
CAGCCCTGAG GGCAGCTGGT TGTCCATTTT TATGGTTATT TCTTGATTGT ATGCTAAACA
AGGGATGGAT TATTCATGCC TCCCCTTTTT AGACCATAGA GGGTAACTTC CTGACATTGT
CATGGCATTC TTTTTTTTTT TTGACAGAGT CTCACTGTGT TGTCCAGGCT GGAGTGCGGT
AAAGCAATCT TGGCTCACTG CAACCTCTGC CTCCTGGGTT TAAATGATTC TCCTGCCTCA
GCCTCCCAAG TAGCTGGGAT TACAGGCATG ACCCACCATG CCCAGCCAAT TTTTGTATTT

TTAGTTGAAA CAGGGTTTTG CCATGTTGGC CAGGCTGGTC TTGAACTCCT GACTTCAGGT
GATCTGCCTA CCTTGGCCTC CCAAAGTGCT GCAATTACAG ATGTGAGCCA CCAAACCTGG
CCTGTCATGG CATTCTTAAA CTGTCATGGT GCTGGTGGGA GTGTAGCAGT GAGGAAGACC
AGAGGTCACT CTCATCGCCA TCTTGGTTTT GGTGGGTTTT AGCTGGCTTC TTTACTGCAG
CCTGTCTGTT CTATCAGCAA GGTCTTTATG ACCTGAATCT TGTGCTGACC TCCTATCTCA
TTCTATGACT TAGAATGCCT TAACTGTCTG GGAATGCAGC CCAGTAGGTC TCAGCCTCAT
CTTACCCAGC TCCTATTCAA GATGGAGTTG CTCTGGGTCA AACACCTTTG ACAACATCAT
TAAGCCTCAG TTCTCACACT GTTTTTGTTT TGTTTGTGTG TGCAATGATG GGCAAATCTC
TGCCTTATAG GATGGTAGAA AAAAGGAACT TAATATTGTA ATGACTGTTT TGTGCCAGAT
AATTGCTTTA AACTGTATCA GCATCTTATT TAGTCCTGTT AATGATATGA ATGTTATCTT
CATAACAGCT GCCATTTTAT TAAGGACTTA TCAGAGAAAA ACACTGTTCT AAGCACTTGT
TACCCATTAT TACATTGAAT TTTCATAACA ACCCTTTGAG GTAAGCATGA TTATACCCAC
TTTAATAGAA GTGAACTGTA GTTTTGTAAT GTTAGGTTCC TTGCCAAATG TTACACAGAT
AGTAAGTGAT AAAATCATAT GCCCTGAAAT TACATTATGC TGCCAAACTT AAATTTCTTT
TTTATCCTTT ATATTAGTAT ATTCTTAGGT TTAAACAAGA CAACTAGTTA ACACATACTA
GATTTTGTCC ACAGTTCCTG GCTCATAACT CCCATAGCCC TTGTCACTAT CTTTTAANCG
TTGGGGCACA TTAGGCCTCA GAAGTAGGCC TCANGAAAAC AGAGTCTCTC TCTCTCTCTC
TCTCTGATCT TCTTCCGCCC TCCTCTCACC TGCCCAGGGC AGCACTTTAA TCTTCTCCTG
CCTTTCTGAT CTTGGGTCAT AAGACCTTCA TTTCCAAAGA TGTCCTGTGT CATACCCTAA
AGGAAGGAAC ACTGAACAGA GAGAGGCTCA GAAGAATCTG GACAGGCCTT GCTGTGTTTA
CATCATTCCC TTTATGTCCA GTCACATCTC TACATGGTTG TCAGTTGTGC CTATTTGATG
AAGTCCCCAT ATAAGGCTCA CAAGGACAGG GTGCAGAGAG CTTCCAGATA GCTGAACAAG
TGGAAGTTCC TGGAGGGTGG CGTGTTCAGG GAGGGCATGG AAGCTGTGTG CCCCTTCCCC
CATACCTTGC CCTACTCATT TCTTCATCTG TTTCATTTGT AGTATCTTTT ATAATAAACC
ACTAAACATT AGTTAGTATT TCTCTGAGTT CTGTGAGTCA CTCTAGCAAA TTAATTGAAC
CCAAGGAGGG TGTCATAGGA TCCCCNACAT TATAGCTGGT TGGCCAGAAG CACAGGTAAA
CAACCTAGGG CTTTCAATTG GCATGAGAAG TAGGGGGCAG TTTTGTGGGA CGGAGCCCTC
AGCCTGTGAG ATCTGATGCC ATCTCTAAGT ACACAGTGTC AAAACTGGAT TGGAGGACAC
CCAGCTAGTA TTCACTGTGA AATTGGTTGC TTGCTTGATT TGTGGGGAAA AACCCACATG
CATTTGATCA CAGAAGTCTT TTGTGTTGAC AGTTGATAGT GTTCAGTGAG AGAATTAAAA
AAAAATTGAG TTTCTTCTTC AACATACTCT CTCAATGTGA AACCACAGAA ACTATTTCCA
TTCAAAGATG GAAATGGTTT GTTTGCATCT TAGTTTTTAT TTATACATCT TAGAAGAAAT
GTCCAAGCTT TGTTTTTTCT CTCACCCTAT ATATAAAATT ACCTATGAGG CACAGATTTT
TATGATCCTT GATTATATAG ACTTTGTCCA AATTGTGTGT TTTATAGCAT TACTGTAACT
TGTTATAGTA ATCTTTGTGT ATATTATGTC TCTTAACATT GTCTTCCATA TTGTTAATGA
CCATCTCATA TTTATCTCTG TATCATGTAT ATCTTCAACC AATGTGACTG GCTTAGGAGA
AAAAATTAGT GAACAATTAA CTAGTGTTTG TGTAATCTAT ACAATTGTCA AGGTTACAAT
TGCTATTTTT GAAGAAATCG TTGTTGTTTT TCTCTTTGTT TCATCTCAGT TCCATTTTGT
CAAGGATTCC TTTTTTTTTT TTTTTTTTTT TTTTTGAGGC GGAGTCTTGC TCTGTCACCC
GGGCTGGAGT GCAGTGGTGC AATCTCGGCT CGCTGCAAGC TCCACCTCCT GGGTTCATGC
CATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTACAG GCACCCGCCA CCACGCCCAG
CTAATTTTTT TTTGTATTTT TAGTAGGGAC GGGGTTTCAC TGTGTTAGCC AGGATGGTCT
CAATCTCCTG ACCTCGTGAT CCGCCCGCCT CGGCCTCCCA ATGCTGGGAT TACAGGCGTG
AGCCACCGCG CGTGGCCCCT TGTCATGTAT TCTTAACCTG TGTTATATCT AAGAGAAGGT

GTGAGAGGCG GGACCATTTG TGGATAGGTG CAGAGGGCAT GAAGTAGCCC AGAAAGAATC
TTTGCCATTG ACTAAATTTT AGCCTAGTAA AAAACATGTG GTAACCACTG AAAAACTAGA
ACAGTGTGAT ATGACAATGC CCAGTTGAAT AAGAAATTCA AATAGTTTAT AGTAACAAAA
AATAATTTTT ACCACATTGT GACTAGTGCC CTAAAAACAA ATTTGAAGGA GCAGAGAGAG
AGAGGAAGTG AGTTTTCGCT GGGGGCTGGT GAGTGGAGGC TACTTTATGA GCAGTTTTGA
AAATTACAAG TTAGGGAAAA ACTCTTAGGG GAGTTTTAAG TGGTGAGCAA CAGGCACTTG
AGGTTACAGG AGCAGCAGCA TAAAGATATA GAACAGAGAG GCCATTTGGC AGTCTTGGGG
AAACTCCAAG AAGTCGAATG TGTCTAAAGC TGTGGTTCTC TCTCTGTGTG TGTGTGTGTG
TGTGTGTGTG CGCGCGTGTG TTTGTGTGTG TGTGCATGTG TATGTATGTG TGAAATCGCT
AATACAGAGA AGCTGTTAGA ACAATTTTCT CACTTCCCAA AAATGTCCGT TCATTCATGG
GCCTAGGCAA CTCTCCTTTG TGTGTCTTAC AGCCCAACAC TGTCATATAA GGTGTGGTTC
TATAATGAAA CAGTTTTATC TGTTGTTTAC GGAGCTAGTC AGCCTGTGTT ATGCTTGTTA
CTGGTTGAGG GTGTCCAGGT TCTTGGCTTC TTGAACAAAG AATTGGACAA AACTCACAAA
TGAGGCAAGG AAAGAATGAA GCAACAAAAG CGGAGATTTA TTGAAAATGA AAGCACACTC
CACAGGATGG GAGTGGGCCT AAGTAAATGA TTCAAGGCCC TGGTTACAGA ATTTTCTGGG
GTTTCAATAC CGTCTAGAGG TTTCCCATTG GTTACCTGGT GTATGCCTTA TGTAAATGAA
GAGGGTGGAG TTAAGTTACA AAGTCATTTA CTCAGTATAG GCCTTGTGTT AATGGAGAGG
GTGTTACTCC TGGGGGTTGT GGCCCATGTA AACGGAGAGG ATGAAGTGAA GTGACAAAGC
CCTTCGCATT CCTGCCATTG CTGAAGTGTT TCCACTTTAT TTAGTTCTAG GAAGTCAGTG
TGAATTGGCC TTATGTTCCC TGCCTCCAGA ACCTGTTCTC CTGCCTCACA TAGACCCTTT
TTTCCCTTTG CCAACTTCAC TTCTTTTACA GCCACCTCAC TACCAATGTG TCTGTCTCCT
AAGTCAAACA TCAGTGATTC CTCTGTTTTC TCTAAACCCT TCTTATGTCT TCTACTTCTC
ATCTCTTTCT TGGTCAAAAA TCTTTCAAAA ATGAGTAAGA ATGCAGCTAT TCAGGCAAAC
TAAAAATAAC ATCACAGTGA TATACAAAAC CAGTGTCATT TCACAAAGGA AAATTATCAA
TACTAGATCC TGAAAAAGAA ACAGCGAATG AAAGCCATTT ACACAACTCC ATTGTGTAAT
TGACACATTG AATCACTCAT AAAACAGGTG CTCTGGGTCT GAATCTAGAT CCTAGCTAGT
CTGGTAGCTG AAATCATAGA ATTATAGTAG AGTTTAGGAA ATCATCCTCA AAGGAAAGAT
TATATGTTGA TATCAAATGT ATATTTCCTT TCTAGGCCCA AACCTGCACC ACTGACTGCT
GAAATACAGC AAAAGATTTT GCATTTGCCA ACATCTTGGG ACTGGAGAAA TGTTCATGGT
ATCAATTTTG TCAGTCCTGT TCGAAACCAA GGTAAAAAAA TAAGCCTAAG TTTTTTGTTA
ATTTGTTTGG AACTATTTAT TGAACAGTTG CTCTGTGTGA TGGATTTCGG GGATACCTAG
ATGGAATGGG CATGATCCCC TTTTTACAGA AATAGAAAAT AGGTGGCCTA TGAATTATTC
TTTCCTTTTA TATCCATGAC AAACTTTAGT AAAAAATTTT CTTTTCTACT GAGCTTAGCA
TTTATTCAGT ACTCTTCTCA ATATATTTTC CAGGTAGCTA GTGACAATCA GAGTGATATG
TAAGACAAAC TCATTTGTCC TCCTAGTAGG AAAGATTCTA GTAGAAGCAA AGAATTGTGT
ACCATTCTGC AAGTGGTTTG TTGGAATCTT TCTTTGATAC CTGTTCCTGT ATTCCCCCAC
CCCCATTAAT TTAGTCATTA ATTACTACAT GAACCCGTAA AATAAATCCT TAAATTATTT
TCCTGGTAGA TTTTTTGAGC TGTGTAAGGA CCTTTCAATT CACTTTACAT TAGAATACGA
TATGTGATGG TAAGTATTAA CCCAGCTTCC TGAGTGATGG CGTGAGGGCG GAACCCAGGT
TCATGAAAAG TATTCCTATT ATTGAATTTT ACCAGTTATT TTCAGAAGTT TGATACAATG
TGAGTTGATT TCATAACATG TCCTCTAACT GCACTTGATA GCAAATTATC TTGTTATCCT
GAGTTGTAGC CAACAATGAC TTGGAGAATC TATGCAATAC TCAGTTTTAT TACTTCTAAG
CTCATTTTGA AGATAATACT ACCCATGGCT GATTTGTTAC TATAAAATAG GTTTAGTATT
TGCTGTCTGG AAACATTCTG ATTAGTGTCT CCTGGGAGGA TTATAAATTT ATAGTACCCA

AAGANTAAAC NTGTTGTTTC CCTTTCCTAA ACTTTTAGTG NATAATNCAG TCTCTGCCGT
GTCTCATTTT CATCACTTGC CCTCCANAGC TCCCATCTCA CTGAATTCTT GCAGTGTTCT
GAATGTTGAG AGCCCCAANG TGGGTCTTAT AACAGCCAGT CAGCAACATT TCTGTTTTTC
ATCTGACACC AAGGGTCTCG TCTCTTTGCT TTTCTACCAG TTATTCTGGG CTCTTCAGCT
CTAAAGAAAG TATAGGTCCT GAAATCTTTC CCTACCTTCT CAATTTCCTG GGGAGGGCTT
CTTTGGAAAG TGGGATTGGA AATAAGATAA ATTTGAAGAT AATTGAGAAA TGAATGGAAA
GTGAAATTGA AGGGTCCATG TTAAGAGATT GCAAGTTATG CTATCACCAA ATAGATTTTT
TGTGCCTGAG AGGATTAATT CATAGTGCAT ATTATGTGTT GACTTTATCA TTGAGGTCCT
GGCACATGAT AGCATTGGCA TGATATAATT TGAGCTACTG ATACTATAGT GTTGCTTCTG
GTGTTGTTAC CAGAACATCT AAAATATATT AGGATTTTTT TAATGGCAGA GGAAATGAAG
ACTAATATGA CATAGTCCTT GTCCTAGTGA TTTACAGTTT AGCAAGACAT ACAAGCAAAA
CATTAAAGTA AAGCATGATA ACTACTTTAA TAAAGCATTT TTTAAATTCA TTGGTAACAT
AAGAGAAGGT AGAAGAGTTT AGCAAACCCT TCCCAAAAGA AATGATTGAC AAGTTATATG
AGATAATAAT TCAGGGAAAG GAAATTCGAC TTTCTAAAGC CAAATTATTT GACATTGGTT
TTCATATAGC TTGGTAAAAG CTGTTATTTT CTCCCATGTT CTTTATTCTT TGACTGTTAT
AAGTATGATT TGTACAGAGA AATGGCAATT TCAAAACAGA GGGCTTTGAT GGATTAATTG
CTTTGAATTG ATCCCTCATC TACAGTATCT TGTCAGGTAC TTGAGAAAAT AATGTACTTA
AAGTTTCCTC TTTTGACTTT CTTTTGGTAT TCTATACTGT AAGTTGGGGA AAAAAGTATT
TTCTCTTCCT GCTAATTGGG CTACTTGAAA ATTCCCACCA ACTTTGCCAA TACCAGTGTT
CTGTATAACC CAGAATTCAG AATTAGCTGC AATTAAGGGA ATTCACAGCT TTTCTGTAGT
CAGAGAGCAA ATTGAAATTA AAGAAAAAAG AAATAGTGGG AGGACAAATG AGGTTTTACC
TTTACACTTG AAAACAGATT TAAGAACAAG CCTTATACCT AGATTTATTA ATACTTTGAG
GTATGAGAGG GAAGAGAAGC TTAGAAATAC GGCAGAATGG GCTTTCTTTG TTCTTCTCCC
AGCTATGCTG TTTTTATTTA TTATTGTATT TTTAAGAGAG AAGGGAAGTG TCTCTCCTGG
GTCACATTAA TTAGGAAATA CAGAGTGTTT TCATAATGCG TAAAGTCTAG TCCATTTAAG
TCTTGTTTCA AAATGCTATT TATATTATTT GAGCAGGAAG GCAGAGACCT TAAACTGCTC
CACCCAATTC ATTTTACACA AAAGATTAAA AAGAAAAAAA CAGTGTCAAA AGGTCAAGTG
CCCAGTGTTG CACAACTAAT GATAGGCAAA ACAAGAGAGG AGGGGAAAAA AAAGAATCCT
GACTCCCGAC TCCTAGTTCA GTGTGTTCTT CACTGTTAGA AGGTGCTGCT GAACATAGTT
ATACCATATG AAGATCACAC TATCTATTTG AGATTGTAGA AATTTGATTA CTGCAGAGCT
CTGGCTGGGC TAGATTCACT TCTTATTCTT TATTGCATTG CAGTATTCTT AAGAGATAAA
TGGCTCTTTT AGAATCAGAT TGGCCTTGGC TGTTGAAATG AGGCATTAAT TACCTTGGTA
GCTGACACAT TCTTACAGGT CAGGGGCTTG ATGAAGTTTA TCTTCTTCCC TTGTTCCTGT
GATTGCTCTG TAAATAGAAC ACATTCAGAG CCCTTGAATG CACTTGCAGC CTGTGCCTCC
CACAGTGATC GATGGTCAGA TAATGGGAGT TTAATGACCA GTACTGAGAG AGATTATTTC
CATGGCTGCT ATGGGCCAAG AAAGCTGGGT GGTCAGAAAG GGACCTTTTC CAGACTCTCC
TGGGTTGTGT TATTTCTTTG ACATTGGTTT CCTTTCATGA GGCGCCAAGG TGTATTTGTA
AGTTGTCCAG TGTTGCACAG CTAATGAGGG GCAAAACAAG AAAGGAGGAA AAAAAAAGTA
TCCTCACTCC TAGTTCAGTG TGTTCTTCAC TGTTGGAAGG TACTGCTGAA CACAGTTATA
TGTTCAGTGT ACAATGTACA ATGTTCTTTG ACATTCCGTC ATTTGAAGGA TGGGTCCTAT
GGATAGTATC ATCTGCATGG TTTTGAAAAC AAAAGATATA AGCTAATTTT GCCCTGTCTA
GTGACTACGA GACAGGGAGA GAAAATCTGA ATATTTGTTA AAGTAGACAC AGACCCATAA
ATTGAAAAGG ACACTAATCC TGCCTTAGGA GACAGTAAGG CACTTGTCCC TGCTATCTAT
TAGCTGTGTG TCCTTGGACA GCTCATTGCT TCTTTCTGAG CCACCGTTAC ACGTTATCTA

TCAAGTGAGC AGGTTGTACA CTAGATGAAT TCACAGGTCC TTCTTCCAAG TGCTTTTCTA
ATCTTCATAA TTTAGATAAT CTCTCAGTAG CAAAACAGTG TACAATATGG TCAATCTGAG
ATTTTTAGGG GGAGAATTTT AGGGAACATC AGAAATGGCA GTAGTTAAAA GGAAATAGGA
CTCACAGGCT GACTTCTCTC TATAACTTCA CATGGTAGAA GGGATGAGAG TTCTCTCTGG
AGCCTTTTGT ATTAGGTCAC TAATCCCAAA GGCCCCACCT CCTAAGACCA TCACTTTTGG
GATTGGGGTT TTAACATGAT TTTAGAGGGA CATAAACCTT CATCTCATTG CATTAGAGAT
ATTTGAAAGC TCAGCTCACG TGTATTCCTC CACAGCTCAC ATGTATTCCT CCATCACCCA
ACCTGATGGC TTTGAACGTT GTAGACATAA ATCCTTTCAT CATTATCAAG AATATTGCCA
AAAGCTTCTC AGAAATTATG AGGGGTTTTT TTAGTTTCTA AAATATTCTC AAAGAAAGTC
CCATGTACTA ATGTTTGCCT TTTGATGAAA AAGGATGAAA TCTTAATGAT TGCCTTAATA
AGCTCAACAA TGCTTGTTAG TTGAGTCTTC TTATTGTGCT GATTCTTATA AACAACAACA
TTCAGTATAA ACATTAATGT ATGTGATTCA CTAAGGTTTT TGCATGATTC TCTGTGAGGC
TTCAGATGTC TCTTGGATTA TGTGTCTTTT TTTCATCGCC AGCATCCTGT GGCAGCTGCT
ACTCATTTGC TTCTATGGGT ATGCTAGAAG CGAGAATCCG TATACTAACC AACAATTCTC
AGACCCCAAT CCTAAGCCCT CAGGAGGTTG TGTCTTGTAG CCAGTATGCT CAAGGTAAGT
GTTGCATTTC AGACACCATT TATGAGCTAT TTACCTGTGT GCAGCTGGCT GTTGTTGGCA
AAGGCAAAAG GATGATGCAG TAGAGAGAGC GCAGTGTCTA TAGTCAGAAA ATCTGAGTGC
AAGTCTGGCC CTATCACTTA TTAATGGATG ATTGCTCATG GAATTTACTG TACCATCCAG
CAAAATGTCA ATAGTTACTA TATATTGAGT GAGCTCTGCT TGTTACATAT ATTGCCTAAC
AATGCTCAAA ACTCTGAGAA GTAGTAAGTA TAATCCCTAT TTATGGGCGG GGAACAGGAA
CTAAGAAATT TTTCTAATAA TTTGAAGGTC TCACAGCTTT TAGCATTGGA GTTTCACTTC
TAATCATCGT CTCCAAAACC CAACTTTTAT TAAAACTATA CTAACACTGG TTTCTCTCTG
GGAGAATTTT AAAATTCTGT ACTTAGGGCT GGGCACAGTG GCTTATGCCT ATAACCCTAT
CACTTTGGGA GGCTGAGATG GGTGAATCTC TTGAGTCCTA GAGTTTGAGA CCAGCCTGGG
CAACACGGCG AAACCCCTTC TCTATTAAAA ATACAAAAAA TTAGCTGGGC GTGGTGGTGT
GTGCTTGTAG TCCCAGCTAT TCAGGAGGCG GAGGTGTAAG AATCACCTGA GCCCAGGAGG
TCAAGGCTGC AGTGAGCCGA TATCATGCCA CTGCACTCCA GCCTGGGCAA ACGGAGTGAG
GCCCTGTTAT GAAAAAAAAA AAATCTGTAC TTAGGCTTTC AGATCAGGCT GTATGTGATG
TATGTCGAAA ACACAGCTAT AATTGATTGA GGGAGAAACG TTACCATTTT AAAGTTTATG
CTTTCAAGCC CAGATTTGGC CACTAGGAAT TTCCCAGCTC ACTAGTGAAA CTGCTGATGA
GTGATTATTT GCCAGTGAGC CTTTCATTCT TTCTAAAATA TGTACTACTA GTTGTGACTT
GTAGGCTATA GGGGCTATAA TATATCAAGA CAATCTTTAT CCTCATGAAG CTTACAGTTA
AGTAAGAGAT AGAGATTAAA TAATTATAAC AACAGAGTGA AGAACAGTGA AGAAAAAGTA
CAGAGTTATA ATATATATAA TAGGGCCAGG ACTGCATGAG GAAGGTAGGA AAGACATTTC
GGCAAGAGGT TGTCAGGGAA AAGACTTGCT TGAGAAAGAG CCAAGTTGTG GGGTCTGGCT
GCTTAGCAAT GACCATAATA CCTAACTTTT GCTATTTTTA CATGAAGTAA CTAATTTAAC
CCTATGAGGA AAGTACTACT ACCATCTAGA TTTTACAGGT AAGTAAGCAG AGATACAGAG
AAGTTAAACT CTTCACACGG CTTTGGCTTT AAACCTATAT AGGCTTCAGA GCCTCCCCAC
TTAACCACTT TGCCATAGCT ACATCCATAT TAGGTGCTAA GTAGATATCT GTTAAGTAGA
AGGAGGATGA AAGGATAGTT AGCTAGTTGG AAAATGGATG GATGAATGAA GTGATGCTTA
AGCTAAGAAC AACTTTCAGG GGTAACATGC AAAGAATAAT GGAGCAAAGA AGAAAAAATA
GAAAATGGGA TAATCCTTTT TCTACTAAGG GGTAACCATG TGTGTTATTC ATCTTCAGGC
TGTGAAGGCG GCTTCCCATA CCTTATTGCA GGAAAGTACG CCCAAGATTT TGGGCTGGTG
GAAGAAGCTT GCTTCCCCTA CACAGGCACT GATTCTCCAT GCAAAATGAA GGAAGACTGC

TTTCGTTATT ACTCCTCTGA GTACCACTAT GTAGGAGGTT TCTATGGAGG CTGCAATGAA
GCCCTGATGA AGCTTGAGTT GGTCCATCAT GGGCCCATGG CAGTTGCTTT TGAAGTATAT
GATGACTTCC TCCACTACAA AAAGGGGATC TACCACCACA CTGGTCTAAG AGACCCTTTC
AACCCCTTTG AGCTGACTAA TCATGCTGTT CTGCTTGTGG GCTATGGCAC TGACTCAGCC
TCTGGGATGG ATTACTGGAT TGTTAAAAAC AGCTGGGGCA CCGGCTGGGG TGAGAATGGC
TACTTCCGGA TCCGCAGAGG AACTGATGAG TGTGCAATTG AGAGCATAGC AGTGGCAGCC
ACACCAATTC CTAAATTGTA GGGTATGCCT TCCAGTATTT CATAATGATC TGCATCAGTT
GTAAAGGGGA ATTGGTATAT TCACAGACTG TAGACTTTCA GCAGCAATCT CAGAAGCTTA
CAAATAGATT TCCATGAAGA TATTTGTCTT CAGAATTAAA ACTGCCCTTA ATTTTAATAT
ACCTTTCAAT CGGCCACTGG CCATTTTTTT CTAAGTATTC AATTAAGTGG GAATTTTCTG
GAAGATGGTC AGCTATGAAG TAATAGAGTT TGCTTAATCA TTTGTAATTC AAACATGCTA
TATTTTTTAA AATCAATGTG AAAACATAGA CTTATTTTTA AATTGTACCA ATCACAAGAA
AATAATGGCA ATAATTATCA AAACTTTTAA AATAGATGCT CATATTTTTA AAATAAAGTT
TTAAAA

The corresponding cDNA sequence for CTSC is provided below (SEQ ID NO: 2):

```
   1 aattcttcac ctcttttctc agctccctgc agcatgggtg ctgggccctc cttgctgctc
  61 gccgccctcc tgctgcttct ctccggcgac ggcgccgtgc gctgcgacac acctgccaac
 121 tgcacctatc ttgacctgct gggcacctgg gtcttccagg tgggctccag cggttcccag
 181 cgcgatgtca actgctcggt tatgggacca caagaaaaaa aagtagtggt gtaccttcag
 241 aagctggata cagcatatga tgaccttggc aattctggcc atttcaccat catttacaac
 301 caaggctttg agattgtgtt gaatgactac aagtggtttg ccttttttaa gtataaagaa
 361 gagggcagca aggtgaccac ttactgcaac gagacaatga ctgggtgggt gcatgatgtg
 421 ttgggccgga actgggcttg tttcaccgga aagaaggtgg gaactgcctc tgagaatgtg
 481 tatgtcaaca cagcacacct taagaattct caggaaaagt attctaatag gctctacaag
 541 tatgatcaca actttgtgaa agctatcaat gccattcaga agtcttggac tgcaactaca
 601 tacatggaat atgagactct taccctggga gatatgatta ggagaagtgg tggccacagt
 661 cgaaaaatcc caaggcccaa acctgcacca ctgactgctg aaatacagca aaagattttg
 721 catttgccaa catcttggga ctggagaaat gttcatggta tcaattttgt cagtcctgtt
 781 cgaaaccaag catcctgtgg cagctgctac tcatttgctt ctatgggtat gctagaagcg
 841 agaatccgta tactaaccaa caattctcag accccaatcc taagccctca ggaggttgtg
 901 tcttgtagcc agtatgctca aggctgtgaa ggcggcttcc catacctat tgcaggaaag
 961 tacgcccaag attttgggct ggtggaagaa gcttgcttcc cctacacagg cactgattct
1021 ccatgcaaaa tgaaggaaga ctgctttcgt tattactcct ctgagtacca ctatgtagga
1081 ggtttctatg gaggctgcaa tgaagccctg atgaagcttg agttggtcca tcatgggccc
1141 atggcagttg cttttgaagt atatgatgac ttcctccact acaaaaaggg gatctaccac
1201 cacactggtc taagagaccc tttcaacccc tttgagctga ctaatcatgc tgttctgctt
1261 gtgggctatg gcactgactc agcctctggg atggattact ggattgttaa aaacagctgg
1321 ggcaccggct ggggtgagaa tggctacttc cggatccgca gaggaactga tgagtgtgca
1381 attgagagca tagcagtggc agccacacca attcctaaat tgtagggtat gccttccagt
1441 atttcataat gatctgcatc agttgtaaag gggaattggt atattcacag actgtagact
```

1501 ttcagcagca atcttcagaag cttacaaata gatttccatg aagatatttg tcttcagaat
1561 taaaactgcc cttaatttta atataccttt caatcggcca ctggccattt ttttctaagt
1621 attcaattaa gtgggaattt tctggaagat ggtcagctat gaagtaatag agtttgctta
1681 atcatttgta attcaaacat gctatatttt ttaaaatcaa tgtgaaaaca tagacttatt
1741 tttaaattgt accaatcaca agaaaataat ggcaataatt atcaaaactt ttaaaataga
1801 tgctcatatt tttaaaataa agttttaaaa ataactgc

The wild type CTSC protein sequence is set forth below as SEQ ID NO: 3:

MGAGPSLLLAALLLLLSGDGAVRCDTPANCTYLDLLGTWVFQVG
SSGSQRDVNCSVMGPQEKKVVVYLQKLDTAYDDLGNSGHFTIIYNQGFEIVLNDYKWF
AFFKYKEEGSKVTTYCNETMTGWVHDVLGRNWACFTGKKVGTASENVYVNTAHLKNSQ
EKYSNRLYKYDHNFVKAINAIQKSWTATTYMEYETLTLGDMIRRSGGHSRKIPRPKPA
PLTAEIQQKILHLPTSWDWRNVHGINFVSPVRNQASCGSCYSFASMGMLEARIRILTN
NSQTPILSPQEVVSCSQYAQGCEGGFPYLIAGKYAQDFGLVEEACFPYTGTDSPCKMK
EDCFRYYSSEYHYVGGFYGGCNEALMKLELVHHGPMAVAFEVYDDFLHYKKGIYHHTG
LRDPFNPFELTNHAVLLVGYGTDSASGMDYWIVKNSWGTGWGENGYFRIRRGTDECAI
ESIAVAATPIPKL.

Papillon Lefevre syndrome (PLS) is an autosomal recessive disorder characterized by palmoplantar hyperkeratosis and severe early onset periodontitis that results in the premature loss of the primary and secondary dentitions. The 46 kb *CTSC* gene consists of 7 exons and is mutated in PLS patients. Sequence analysis of *CTSC* from PLS affected individuals from thirty-two Turkish families identified four different mutations. An exon 6 nonsense mutation (856C->T) introduces a premature stop codon at amino acid 286. Three exon 2 mutations were identified including a single nucleotide deletion (1047delA) of codon 349 introducing a frameshift and premature termination codon, a two base pair deletion (1028-1029delCT) that results in introduction of a stop codon at amino acid 343, and a G->A substitution in codon 429 (1286G->A) introducing a premature termination codon. All PLS affected individuals examined were homozygous for cathepsin C mutations inherited from a common ancestor. Parents and siblings heterozygous for cathepsin C mutations do not show either the palmoplantar hyperkeratosis or severe early onset periodontitis characteristic of PLS. In

addition to the 5 families described above, Table I summarizes CTSC mutations identified in 27 other families presenting with symptoms of PLS.

Haim-Munk syndrome is a rare condition associated with congenital palmoplantar keratosis, pes planus, onychogyrphosis, periodontosis, arachnodactyly and acroosteolysis. In an additional embodiment of the invention, a mutation in cathepsin C causing Haim-Munk Syndrome has been identified. It appears that substitution of an A for a G at CTSC nucleotide position 857 in Exon 6 is responsible for this syndrome in patients.

Based on the data presented herein, it appears that additional mutations or functional polymorphisms are associated with other pathological conditions, including, but not limited to prepubertal periodontitis (PPP), early onset periodontal disease or other forms of gum disease. For example, as shown herein, PPP is caused a substitution of a G for an A at position 1040 in the CTSC coding sequence. Thus, the invention also provides methods for screening the CTSC gene for alterations associated with these disease states.

I. Preparation of Altered Human CTSC-Encoding Nucleic Acid Molecules, CTSC Proteins, and Antibodies Thereto

A. Nucleic Acid Molecules

Nucleic acid molecules encoding the human CTSC proteins of the invention may be prepared by two general methods: (1) synthesis from appropriate nucleotide triphosphates, or (2) isolation from biological sources. Both methods utilize protocols well known in the art. The availability of nucleotide sequence information, such as a DNA having the sequence of SEQ ID NOS:1-2 enables preparation of an isolated nucleic acid molecule of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the

phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct may be purified according to methods known in the art, such as high performance liquid chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current oligonucleotide synthetic methods. Thus, for example, a 4.7 kb double-stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire 4.7 kb double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an appropriate vector.

Nucleic acid sequences encoding the altered human CTSC proteins of the invention may be isolated from appropriate biological sources using methods known in the art. In a preferred embodiment, a cDNA clone is isolated from a cDNA expression library of human origin. In an alternative embodiment, utilizing the sequence information provided by the cDNA sequence, human genomic clones encoding altered CTSC proteins may be isolated.

Table 1 sets forth several different mutations associated with particular PPKs and PPP. Altered CTSC-specific probes for identifying such sequences may be between 15 and 40 nucleotides in length. For probes longer than those shown above, the additional contiguous nucleotides are provided within SEQ ID NOS:1 and 2.

Additionally, cDNA or genomic clones having homology with human CTSC may be isolated from other species using oligonucleotide probes corresponding to predetermined sequences within the human CTSC encoding

nucleic acids.

In accordance with the present invention, nucleic acids having the appropriate level of sequence homology with the protein coding region of SEQ ID NO:1 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989), using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing the solution every 30 minutes.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, clones are maintained in a plasmid cloning/expression vector, such as pBluescript (Stratagene, La Jolla, CA), which is propagated in a suitable *E. coli* host cell.

Altered CTSC-encoding nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the DNA having SEQ ID NO:1. Also contemplated in the scope of the present invention are oligonucleotide probes which specifically hybridize with the mutated CTSC genes of the invention while not hybridizing with the wild type sequence under high stringency conditions. Primers

capable of specifically amplifying the altered CTSC encoding nucleic acids described herein are also contemplated herein. As mentioned previously, such oligonucleotides are useful as probes and primers for detecting, isolating or amplifying altered CTSC genes.

Antisense nucleic acid molecules may be targeted to translation initiation sites and/or splice sites to inhibit the expression of the CTSC gene or production of the CTSC protein of the invention. Such antisense molecules are typically between 15 and 30 nucleotides in length and often span the translational start site of CTSC encoding mRNA molecules.

Alternatively, antisense constructs may be generated which contain the entire CTSC cDNA in reverse orientation. Such antisense constructs are easily prepared by one of ordinary skill in the art.

It will be appreciated by persons skilled in the art that variants (e.g., allelic variants) of CTSC sequences exist in the human population, and must be taken into account when designing and/or utilizing oligonucleotides of the invention. Accordingly, it is within the scope of the present invention to encompass such variants, with respect to the CTSC sequences disclosed herein or the oligonucleotides targeted to specific locations on the respective genes or RNA transcripts. Accordingly, the term "natural allelic variants" is used herein to refer to various specific nucleotide sequences of the invention and variants thereof that would occur in a human population. The usage of different wobble codons and genetic polymorphisms which give rise to conservative or neutral amino acid substitutions in the encoded protein are examples of such variants. Such variants would not demonstrate altered CTSC activity. Additionally, the term "substantially complementary" refers to oligonucleotide sequences that may not be perfectly matched to a target sequence, but such mismatches do not

materially affect the ability of the oligonucleotide to hybridize with its target sequence under the conditions described.

B. Proteins

Full-length, altered, human CTSC proteins of the present invention may be prepared in a variety of ways, according to known methods. The proteins may be purified from appropriate sources, e.g., transformed bacterial or animal cultured cells or tissues, by immunoaffinity purification. However, this is not a preferred method due to the low amount of protein likely to be present in a given cell type at any time. The availability of nucleic acid molecules encoding CTSC protein enables production of the protein using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such as pSP64 or pSP65 for *in vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocyte lysates. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or Gibco-BRL, Gaithersburg, Maryland.

Alternatively, according to a preferred embodiment, larger quantities of CTSC protein may be produced by expression in a suitable prokaryotic or eukaryotic system. For example, part or all of a DNA molecule, such as a DNA having SEQ ID NOS:1 or 2 containing an alteration set forth in Table 1 may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*. Such vectors comprise the regulatory elements necessary for expression of the DNA in the host cell positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include

promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The human CTSC protein produced by gene expression in a recombinant procaryotic or eukaryotic system may be purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted from the host cell, and readily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein or nickel columns for isolation of recombinant proteins tagged with 6-8 histidine residues at their N-terminus or C-terminus. Alternative tags may comprise the FLAG epitope or the hemagglutinin epitope. Such methods are commonly used by skilled practitioners.

The human CTSC protein of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. For example, such protein may be subjected to amino acid sequence analysis, according to known methods.

The present invention also provides antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward altered human CTSC proteins may be prepared according to standard methods. In a preferred embodiment, monoclonal antibodies are prepared, which react immunospecifically with the various epitopes of the CTSC protein described herein. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies that immunospecifically interact with altered CTSC proteins can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized

for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of anti-CTSC antibodies are described below.

II. DETECTION OF KERATODERMAL DISORDERS/DYSPLASIAS and PERIODONTAL DISEASE-ASSOCIATED MUTATIONS AND DIAGNOSTIC SCREENING ASSAYS THEREFORE

Currently, the most direct method for mutational analysis is DNA sequencing, however it is also the most labor intensive and expensive. It is usually not practical to sequence all potentially relevant regions of every experimental sample. Instead some type of preliminary screening method is commonly used to identify and target for sequencing only those samples that contain mutations. Single stranded conformational polymorphism (SSCP) is a widely used screening method based on mobility differences between single-stranded wild type and mutant sequences on native polyacrylamide gels. Other methods are based on mobility differences in wild type/mutant heteroduplexes (compared to control homoduplexes) on native gels (heteroduplex analysis) or denaturing gels (denaturing gradient gel electrophoresis). Sample preparation is relatively easy in these assays, and conditions for electrophoresis required to generate the often subtle mobility differences that form the basis for identifying the targets that contain mutations are known to those of skill in the art. Another parameter to be considered is the size of the target region being screened. In general, SSCP is used to screen target regions no longer than about 200-300 bases.

Another type of screening technique currently in use is based on cleavage of unpaired bases in

heteroduplexes formed between wild type probes hybridized to experimental targets containing point mutations. The cleavage products are also analyzed by gel electrophoresis, as subfragments generated by cleavage of the probe at a mismatch generally differ significantly in size from full length, uncleaved probe and are easily detected with a standard gel system. Mismatch cleavage has been effected either chemically (osmium tetroxide, hydroxylamine) or with a less toxic, enzymatic alternative, using RNase A. The RNase A cleavage assay has also been used, although much less frequently, to screen for mutations in endogenous mRNA targets for detecting mutations in DNA targets amplified by PCR. A mutation detection rate of over 50% was reported for the original RNase screening method.

A newer method to detect mutations in DNA relies on DNA ligase which covalently joins two adjacent oligonucleotides which are hybridized on a complementary target nucleic acid. The mismatch must occur at the site of ligation. As with other methods that rely on oligonucleotides, salt concentration and temperature at hybridization are crucial. Another consideration is the amount of enzyme added relative to the DNA concentration. In summary, exemplary approaches for detecting alterations in CTSC encoding nucleic acids or polypeptides/proteins include:

a) comparing the sequence of nucleic acid in the sample with the wild-type CTSC nucleic acid sequence to determine whether the sample from the patient contains mutations; or

b) determining the presence, in a sample from a patient, of the polypeptide encoded by the CTSC gene and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or

c) using DNA restriction mapping to compare the restriction pattern produced when a restriction enzyme

cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal CTSC gene or from known mutations thereof; or,

d) using a specific binding member capable of binding to a CTSC nucleic acid sequence (either normal sequence or known mutated sequence), the specific binding member comprising nucleic acid hybridizable with the CTSC sequence, or substances comprising an antibody domain with specificity for a native or mutated CTSC nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labeled so that binding of the specific binding member to its binding partner is detectable; or,

e) using PCR involving one or more primers based on normal or mutated CTSC gene sequence to screen for normal or mutant CTSC gene in a sample from a patient.

A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The skilled person is aware of many other examples and they do not need to be listed here. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a large molecule. In embodiments in which the specific binding pair are nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

In most embodiments for screening for susceptibility alleles, the CTSC nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other

sequences present in the sample. This allows the target CTSC sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

The identification of the CTSC gene and its association with keratodermal disorders/dysplasias and peridontal diseases paves the way for aspects of the present invention to provide the use of materials and methods, such as are disclosed and discussed above, for establishing the presence or absence in a test sample of a variant form of the gene, in particular an allele or variant specifically associated with PLS, HMS or periodontal diseases. This may be for diagnosing a predisposition of an individual to PLS, HMS or periodontal disease. It may be for diagnosing PLS, HMS or periodontal disease in a patient with the disease as being associated with the altered CTSC gene.

This allows for planning of appropriate therapeutic and/or prophylactic measures, permitting stream-lining of diagnosis, treatment and outcome assessments. The approach further stream-lines treatment by targeting those patients most likely to benefit.

According to another aspect of the invention, methods of screening drugs for therapy to identify suitable drugs for restoring CTSC product functions are provided.

The CTSC polypeptide or fragment employed in drug screening assays may either be free in solution, such as gingival crevicular fluid, affixed to a solid support or within a cell. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may

determine, for example, formation of complexes between a CTSC polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a CTSC polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to CTSC polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different, small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with CTSC polypeptide and washed. Bound CTSC polypeptide is then detected by methods well known in the art.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional CTSC gene. These host cell lines or cells are defective at the CTSC polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of CTSC defective cells.

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, (1991) Bio/Technology 9:19-21. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., CTSC polypeptide) or, for example,

of the CTSC-substrate complex, by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) Science 249:527-533). In addition, peptides (e.g., CTSC polypeptide) may be analyzed by an alanine scan (Wells, 1991) Meth. Enzym. 202:390-411. In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved CTSC polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of CTSC polypeptide activity. By virtue of the availability of cloned CTSC sequences, sufficient amounts of the CTSC polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the CTSC protein sequence

provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

III Therapeutics

A. Pharmaceuticals and Peptide Therapies

The discovery that mutations in the CTSC gene give rise to PLS, HMS, and periodontal disease facilitates the development of pharmaceutical compositions useful for treatment and diagnosis of these syndromes and conditions. These compositions may comprise, in addition to one of the above substances, a pharmaceutcally acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Whether it is a polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual.

B. Methods of Gene Therapy

As a further alternative, the nucleic acid encoding the authentic biologically active CTSC polypeptide could be used in a method of gene therapy, to treat a patient who is unable to synthesize the active "normal" polypeptide or unable to synthesize it at the normal

level, thereby providing the effect elicited by wild-type CTSC and suppressing the occurrence of "abnormal" CTSC associated with keratodermal disorders and dysplasias.

Vectors, such as viral vectors have been used in the prior art to introduce genes into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transformation can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The transfected nucleic acid may be permanently incorporated into the genome of each of the targeted cells, providing long lasting effect, or alternatively the treatment may have to be repeated periodically.

A variety of vectors, both viral vectors and plasmid vectors are known in the art, see US Patent No. 5,252,479 and WO 93/07282. In particular, a number of viruses have been used as gene transfer vectors, including papovaviruses, such as SV40, vaccinia virus, herpes viruses including HSV and EBV, and retroviruses. Many gene therapy protocols in the prior art have employed disabled murine retroviruses.

Gene transfer techniques which selectively target the CTSC nucleic acid to oral tissues are preferred. Examples of this include receptor-mediated gene transfer, in which the nucleic acid is linked to a protein ligand via polylysine, with the ligand being specific for a receptor present on the surface of the target cells.

The following methods are provided to facilitate the practice of the present invention.

Family material and clinical diagnosis.

Five Turkish families were described previously [8]. All available family members provided consent for

the study and were clinically examined. A diagnosis of PLS was made in individuals with severe early onset periodontitis and the clinical appearance of hyperkeratosis on the palmar and plantar surfaces. All affected individuals also had hyperkeratosis on the knees. DNA was isolated from peripheral blood samples from all available members from these nuclear families using standard techniques (Qiamp Blood Kit, Qiagen).

RNA Isolation, Amplification, and Tissue Expression Analysis.

Total RNA was generated from fresh tissue samples (gingiva, palm, sole, knee) using TRIZOL reagent (Molecular Research Center, Inc.; Cincinnati, OH) according to the manufacturer's protocol. To determine if cathepsin C was expressed in a given tissue, single-tube RT-PCR was carried out using the Access RT-PCR System (Promega; Madison, WI), following the manufacturer's protocol. A portion of each reaction was visualized following agarose gel electrophoresis in the presence of ethidium bromide. Amplification primers located within exon 6 **F** 5'-AGGAGGTTGTGTCTTGTAGCC-3'(nt. 857-877;SEQ ID NO: 4) and exon 7 **R** 5'-AGTGCCTGTGTAGGGGAAGC-3' (NT 981-962; SEQ ID NO: 5) produce an amplicon of 123 base pairs from cDNA. A standard PCR protocol was followed with an annealing temperature of 65°C.

GenBank accession numbers.

Full-length cDNA of CTSC (NM-001814) and full-length genomic DNA of CTSC contained within a BAC vector, Genbank accession number (AC011088). See SEQ ID NO: 1.

Cathepsin C Activity Assay

In unafffected non-carriers, cathepsin C activity ranges from 600-1200 µmol/min/ mg. As carriers of a

cathepsin C mutation do not have clinical manifestations, measurement of cathepsin C enzymatic activity can be used to determine whether at-risk family members are carriers. Enzymatic activity can also be used to determine if individuals marrying into a family are carriers. Carriers typically have approximately 50% of normal enzyme activity. Determination of enzymatic activity can also be used to establish a diagnosis of PLS when mutational studies of cathepsin C have been negative. This is important in assuring that a diagnosis of PLS has been properly given to an individual with clinical symptoms suggestive of PLS.

Viable leukocyte pellets are obtained from lithium heparinized whole blood by mixing blood with 3 volumes of 3% dextran in normal saline, and allowing the red cells to settle for 45 min at room temperature. Cells are pelleted by centrifugation at 1500 rpm for 5 min at 4*C. After washing and removal of contaminating red cells, leukocyte pellets are resuspended in dH20 and sonicated on ice for 5 sec each for total of 6 blasts using a Sonic 300 Dismembrator. Protein concentration is determined by the Lowry method.

Enzymatic activity is determined by measuring hydrolysis of the synthetic substrate glycyl-L-arginine-7-amido-4-methylcoumarin at a final concentration of 5 mM using a modified method. All reactions are performed in duplicate. Twenty µl of leukocyte lysate are added to 200 µl of Na_3PO_4 buffer (0.1M, pH 6.5) in a 96 well plate and then substrate added. Reactions are allowed to proceed for 1 hr at room temperature at which time 10 µl of glycine-NaOH buffer (0.5M, pH 9.8) is added to stop the reaction. Fluorescence is determined using a Perkin-Elmer LS50B luminescence spectrometer at 370-nm excitation and 460-nm emission. The amount of NHMec released is determined by generating a standard curve using NHMec. Cathepsin C activity is reported as µmol

NHMec released per min per mg protein.

Sequencing and mutation analysis.

PCR primers were designed to cover the entire cathepsin C gene in overlapping fragments, from 955 nucleotides 5' to the start codon to 240 nucleotides 3' to the termination codon using cathepsin C (DPP-I) sequence data (Accession # U79415; SEQ ID NO: 1). The PCR products were prepared for sequencing by excising the bands from the agarose gel and extracting the fragments using a Qiagen Gel Clean-up Kit. The sense and antisense strand of each PCR product were directly sequenced on an ABI Prism 310 Genetic Analyzer (Perkin-Elmer) using four dye terminator chemistry. Approximately 1-3 ng of purified product and 3.2 pmol primer were added to premixed reagents from the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit, FS (Perkin-Elmer) and underwent a cycle sequencing reaction in a GeneAmp PCR System 9700 (Perkin Elmer). The linear amplification started with a 10 s denaturation at 96°C, 5 s annealing at 50°C and 4 min extension at 60°C. The fluorescently labeled sequencing products were separated from residual reaction reagents using a Centri-Sep spin column (Princeton Separations, Aldelphia NJ) and electrophoresed on POP6 capillary at 1500 V for 30 min. Sequencing data were automatically collected and analyzed by the ABI Prism 310 software.

Table A:

Primers Used to Determine Genomic Organization of Cathepsin C

Region	Primer Sequences
Intron 1	F: 5'- TGTCAACTGCTCGGTTATGGGTAA-3' (#6) R: 5'- TCGAGCTTCTCTTCGTACACCACT-3' (#7)
Intron 2	F: 5'- TGACTACAAGTGGTTTGCCTTTTT-3' (#8) R: 5' TGCTGCCCTCTTCTTTATACTGC-3' (#9)
Intron 3	F: 5'- GCCTCTGAGAATGTGTATGTCAAC-3' (#10) R: 5' CCTGCCCCAAAAATGAGATA-3" (#11)

Intron 4	F: 5' TCGAAAAATCCCAAGGTAATC-3' (#12) R: 5' GGGCCTAGAAAGGAAATATACATT-3' (#13)
Intron 5	F: 5' AATTTGTTCGGAACTATTTATTGA-3' (#14) R: 5' TCGCTTCTAGCATACCCATA-3' (#15)

Exon	Primer Sequences
1	F: 5'-GGCGATCAGACTGGCACACTAGAA -3' (#16) R: 5'-CTTACCCATAACCGAGCAGTTGAC-3' (#17)
2	F: 5'- GCAGACTGTGCTCAAACTGGGTAG-3' (#18) R: 5'-TCTACTAATCAGAAGAGGTTTCAG-3' (#19)
3	F: 5'-GGCACATTTACTGTGAATGAGAG-3' (#20) R: 5'-GTCTCATTTGTAGCAACTCAC-3' (#21)
4	F: 5'-CCACTTTCCACTTAGGCACAG-3' (#22) R: 5'-AGGATGGTATTCAGCATTCATA-3' (#23)
5	F: 5'-ATCCTAGCTAGTCTGGTAGCTGAA-3' (#24) R: 5'-TCTAGGTATCCCCGAAATCCATCA-3' (#25)
6	F: 5'-GATTCTCTGTGAGGCTTCAGATGT-3' (#26) R: 5'-GCCAACAACAGCCAGCTGCACACA-3' (#27)
7	F: 5'-TCCCCACTTAACCACTTTGC-3' (#28) R: 5'-ACTTCATAGCTGACCATCTTCC-3' (#29)

Primers for cDNA templates:

F:5'-GCCGCCCTCCTGCTGCTTCT-3' (#30)

R: 5'-GGCTTAGGATTGGGGTCTGA-3'(#31)

We analyzed raw sequence data, generated consensus sequences, and produced nucleotide/amino acid alignments (DNASIS V2.6 for Windows, Hitachi Software Engineering Co., Ltd.). Mutations were detected by creating nucleotide/amino acid alignments of reported cathepsin C sequence data versus affected PLS patients sequence data using the Higgins-Sharpe UPGMA. Numbers in parentheses are SEQ ID NOS:.

Example I

PLS families

Parents of most families were consanguineous. Linkage studies localized a PLS gene in these five families to chromosome 11q14 [8]. Most affected individuals were homozygous for SSTR markers within the PLS candidate interval on chromosome 11q14, consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent"). Four different haplotypes for short sequence tandem repeat (SSTR) markers spanning the

critical region were identified (Fig. 2), consistent with four independent mutations in the gene responsible for PLS.

Analysis of cathepsin C

Using RT-PCR, we found cathepsin C is normally expressed in epithelium from palms, soles, knees and keratinized oral gingiva from unaffected individuals (data not shown). The cathepsin C gene spans approximately 46 kb and consists of 7 exons. Sequence analysis of exonic, intronic and the 5'regulatory regions of the cathepsin C gene revealed PLS affected individuals from these families were homozygous for CTSC mutations that significantly altered the cathepsin C open reading frame.

Exon 6: Two affected individuals from one family were found to have an exon 6 nonsense mutation (856C->T) which introduces a stop codon at amino acid 286 (Fig. 3).

Exon 7: Three different exon 7 mutations were detected (Fig. 4). A deletion of a single nucleotide (1047delA) of codon 349 was found that introduced a frame shift and an early termination codon (TGA) 27 bases downstream. This mutation would result in a mutated protein of 358 amino acids, compared to the normal (wild type) 463 amino acids. A deletion of 2 bases of codon 343 (1028-1029delCT) resulting in the introduction of an early termination codon (TGA), and a truncated protein of 342 amino acids was identified in another family. A G->A substitution in codon 429 (1286G->A) that altered the original TRP codon (TGG) to a terminator codon (TAG) was identified in two affected individuals (#7 and #22) from two additional families. The expected truncated protein is 428 amino acids. Although these families were not known to be related, the fact that affected individuals from these two Turkish families are homozygous for a common cathepsin C

gene mutation and also share a common haplotype for SSTR markers in the PLS candidate interval flanking the cathepsin C gene (D11S931 - D11S1311) suggests that these individuals have inherited the same cathepsin C gene mutation from a common ancestor. Additional mutations were also identified in Exons 2, 3, 4, 5, 6 and 7. Summaries of the mutations identified to date are set forth in Table I and locations are shown in Figure 5.

Papillon Lefevre syndrome is a palmoplantar keratoderma (PPK) with the characteristic clinical features of palmoplantar hyperkeratosis and severe periodontal destruction. The PPKs are a heterogeneous group of diseases all having gross thickening of the palmoplantar skin. Clinically, the finding that distinguishes PLS from other PPKs is severe, early onset periodontal destruction. In affected individuals, the development and eruption of the primary teeth proceed normally, but the eruption of these teeth into the oral cavity is associated with gingival inflammation and subsequent rapid destruction of the periodontium. This form of destructive periodontitis is characteristically unresponsive to traditional periodontal treatment modalities, and consequently, the primary dentition is usually exfoliated prematurely. After exfoliation, the inflammation subsides, and the gingiva resumes a healthy appearance. However, with the eruption of the permanent dentition the process is usually repeated, resulting in the premature exfoliation of the permanent dentition, although the third molars are sometimes spared [5]. Destruction of the alveolar bone in PLS is usually severe, resulting in generalized atrophy of the alveolar ridges, further complicating dental therapy.

Because cathepsin C both localized to the refined PLS candidate interval on chromosome 11q14 and was normally expressed in epithelium from sites affected by PLS it was evaluated as a candidate gene for PLS.

Cathepsin C, or dipeptidyl aminopeptidase I (EC 3.4.14.1), is a lysosomal cysteine protease capable of removing dipeptides from the terminus of protein substrates, but at higher pH it also exhibits dipeptidyl transferase activity [10]. The cathepsin C gene spans approximately 46 kb and consists of 7 exons that encode a 463-amino acid polypeptide with predicted features of the papain family of cysteine proteases [11]. Unlike cathepsin B, H, L, and S, which are small monomeric enzymes, cathepsin C is a large (200 kD) oligomeric protein that consists of four identical subunits, each composed of three different polypeptide chains [12,13]. Expression of cathepsin C (CTSC) is tissue dependent [14]. CTSC is expressed in pituitary gland, spinal cord, aorta, left atrium, right atrium, left ventricle, right ventricle, inter ventricular septum, apex of heart, esophagus, stomach, duodenum, jejunum, ileum, ileocecum, appendix, ascending colon, transverse colon, descending colon, rectum, kidney, skeletal muscle, spleen, thymus, peripheral blood lymphocytes, lymph node, bone marrow, trachea, lung, placenta, bladder, uterus, testis, liver, pancreas, adrenal gland, thyroid gland, salivary gland, mammary gland, fetal heart, fetal kidney, fetal liver, fetal spleen, fetal thymus, and fetal lung.

The CTSC message is also expressed at high levels in immune cells including polymorphonuclear leukocytes and alveolar macrophages and is also expressed at high levels in osteoclasts [11,16]. The pathologic clinical findings of the PLS affected individuals studied here involve severe inflammation and destruction of the gingiva as well as hyperkeratosis of the skin from palmar, plantar and knee sites. In unaffected individuals, cathepsin C is normally expressed in epithelial tissues from sites clinically affected by PLS.

Most parents of the PLS affected individuals in this study are consanguineous. As a result, most PLS

affected individuals in each family are homozygous for cathepsin C mutations inherited from a common ancestor. Yet parents and several siblings who are heterozygous carriers for cathepsin C mutations do not appear to show either the palmoplantar hyperkeratosis or severe early onset periodontitis characteristic of PLS. It appears that the presence of one wild type cathepsin C gene is sufficient to prevent PPK and periodontal destruction in most patients. However, 1 mutation identified to date, (1047 A->G) appears to be associated with the presence of dermatological lesions. A consistent finding in the three linkage reports to date is the lack of a common haplotype among affected individuals from different families. The present report describes 20 different cathepsin C gene mutations associated with PLS. These findings suggest that the CTSC mutations responsible for PLS have arisen independently. Several mutations reported here result in the introduction of premature stop codons. While the W429X mutation encodes a protein shortened by only 35 amino acids, the introduced stop codon is 1 amino acid from the asparagine residue in the active site (Fig. 5). It is likely that such a mutation would cause a conformational alteration that may decrease or abolish activity. Additionally, we have also identified single nucleotide changes that result in missense amino acid changes in several additional PLS affected individuals from other populations, suggesting that structural alterations of cathepsin C may cause PLS.

In addition to the cardinal features of PLS, reports suggest some PLS patients have an increased susceptibility to infections [5]. This generalized increased susceptibility to infection may reflect the more deleterious effects of specific cathepsin C mutations, or may reflect the epigenetic effects of other gene loci. A variety of immunological findings have been reported in PLS affected individuals including

decreased monocyte chemotaxis, decreased neutrophil chemotaxis, impaired neutrophil phagocytosis, altered superoxide production, and decreased blastogenic response, but it has been difficult to extrapolate results of these studies. Consequently, the underlying pathogenesis of PLS has been poorly understood [17]. Immunological findings previously reported for affected individuals from the current families includes decreased PMN chemotaxis and elevated CD11b expression [18,19]. The pathologic clinical findings associated with PLS suggest that cathepsin C is functionally important in the structural growth and development of skin and in susceptibility to periodontal disease. As a lysosomal cysteine proteinase, cathepsin C is important in intracellular degradation of proteins and appears to be a central coordinator for activation of many serine proteinases in immune/inflammatory cells [11]. It is unknown if the profound periodontal disease susceptibility is a consequence of altered integrity of junctional epithelium surrounding the teeth. It is interesting that once teeth are exfoliated, and consequently the junctional epithelium is eliminated, the severe gingival inflammation resolves. A more complete understanding of the functional physiology of cathepsin C carries significant implications for understanding periodontal disease susceptibility. Identification of cathepsin C gene mutations in PLS raises the possibility of creating an animal model to study the development, treatment and prevention of hyperkeratosis and periodontitis.

Classification of the PPKs based upon histological findings, epidermolysis and localization of lesions within the skin (diffuse, linear or focal) has not been helpful in understanding the pathomechanism of disease [20]. Identification of mutations in specific genes has led to development of a revised nosology of these diseases in which PLS is grouped with the palmoplantar

ectodermal dysplasias [2]. In addition to providing insight into both normal as well abnormal epithelial growth and development, identification of mutations in cathepsin C associated with PLS will contribute to the overall nosology of the PPKs.

EXAMPLE II

CTSC Mutation in Haim-Munk Syndrome

Of the many palmoplantar keratoderma (PPK) conditions, only Papillon Lefevre syndrome (PLS) and Haim Munk syndrome (HMS) are associated with premature periodontal destruction. Although both PLS and HMS share the cardinal features of PPK and severe periodontitis, a number of additional findings are reported in HMS including arachnodactyly, acroosteolysis, atrophic changes of the nails, and a radiographic deformity of the fingers. While PLS cases have been identified throughout the world, HMS has only been described among descendents of a religious isolate originally from Cochin, India. Parental consanguinity is a characteristic of many cases of both conditions. Although autosomal recessive transmission of PLS is evident, a more "complex" autosomal recessive pattern of inheritance with phenotypic influences from a closely linked modifying locus has been hypothesized for HMS. As set forth in Example I, mutations of the cathepsin C gene have been identified as the underlying genetic defect in PLS. To determine if a cathepsin C mutation is also responsible for HMS, we sequenced the gene in affected and unaffected individuals from families with HMS. Here we report identification a mutation of cathepsin C (exon 6, 857A->G) that changes a highly conserved amino acid in the cathpesin C peptide. This mutation segregates with HMS in four nuclear families. Additionally, the existence of a shared common haplotype for genetic loci flanking the cathepsin C gene suggests that affected individuals descended from the Cochin

isolate are homozygous for a mutation inherited "identical by descent" from a common ancestor. This finding supports simple autosomal recessive inheritance for HMS in these families. As described above, we also report a mutation of the same exon 6 CTSC codon (856C->T) in a Turkish family with classic PLS. These findings provide evidence that PLS and HMS are allelic variants of cathepsin C gene mutations.

In addition to congenital palmoplantar keratosis and progressive early onset periodontal destruction, other clinical findings shared by these individuals included recurrent pyogenic skin infections, acroosteolysis, atrophic changes of the nails, arachnodactyly, and a peculiar radiographic deformity of the fingers consisting of tapered pointed phalangeal ends and a clawlike volar curve (Figures 6A and 6B). Subsequently pes planus was reported to be associated with the syndrome [24]. This was the first reported association of these clinical findings, and the condition became known as Haim Munk syndrome, or keratosis palmoplantaris with periodontopathia and onychogryposis (HMS1; MIM245010)[22]. Although the palmoplantar findings and severe periodontitis were suggestive of the Papillon-Lefevre syndrome (PLS; MIM245000)[3], the association of other clinical features, particularly nail deformities and arachnodactyly, argued that HMS was a distinct disorder. In contrast to PLS, the skin manifestations in HMS were reported to be more severe and extensive. In addition to a marked palmoplantar keratosis (Figure 6C, 6D), affected individuals had scaly erythematous and circumscribed patches on the elbows, knees, forearms, shins and dorsum of the hands. While the periodontium in HMS was reported to be less severely affected than in PLS, gingival inflammation and alveolar bone destruction are present and severe (Figure 6E, 6F). In a subsequent genetic study of this extended family, Hacham-Zadeh and

coworkers [25] concluded that the syndrome might not behave as a simple autosomal recessive trait. Based upon their estimate of the disease allele frequency in this population (0.1), the absence of the condition in other kindreds of the Cochin isolate, and an inability to document consanguinity for many of the parents of affected individuals, they hypothesized that a "complex" autosomal recessive inheritance pattern with a closely linked dominant modifier locus may be responsible for the HMS phenotype.

HMS families

Pedigrees of the reported familial relationships for the Cochin descendents are shown in Figure 7A. Descendents of the Cochin isolate studied include sibships 2, 3, 4 and 5 in the kindred pedigree originally described by Hacham-Zadeh and coworkers [25].

HMS family genotyping results

All HMS affected individuals from the Cochin kindred were found to be homozygous for all three polymorphic DNA loci (D11S1887, D11S1780 and D11S1367) flanking the cathepsin C locus. Additionally, these individuals shared a common haplotype for these polymorphic markers. These findings are consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent").

Analysis of cathepsin C in HMS

The cathepsin C gene spans approximately 46 kb and consists of 7 exons. Sequence analysis of exonic, intronic and the 5'regulatory regions of the cathepsin C gene revealed that HMS affected individuals from the Cochin kindred were homozygous for a mutation in codon 286 of exon 6 (857A->G) which results in substitution of a conserved glutamine residue at position 286 by an

arginine: Q286R (Figure 8). This glutamine residue is normally completely conserved in wild type cathepsin C from at least five species (data not shown). This was the only sequence change different from the reported, highly conserved, wild type CTSC sequence (GenBank Accession No.: AC011088; SEQ ID NO: 1). All available parents of HMS affected individuals were found to be heterozygous for the mutated (857A->G) allele and the wild type allele. None of the parents or siblings heterozygous for the mutated (857A->G) allele and the wild type allele manifested clinically identifiable characteristics of PPK or had a history of severe, early onset periodontitis.

Restriction Analysis

The Q286R mutation creates an AvaI restriction cleavage site. We utilized this newly created restriction site to develop a rapid test to screen for the Q286R mutation. After amplification of a 465bp fragment encompassing the 3' end of exon 6 using primers: Forward 5'-GTATGCTAGAAGCGAGAATCCGTAT-3' (SEQ ID NO: 32) and Reverse 5'-CCAATGCTAAAACTTGTTGAGACC-3' (SEQ ID NO: 33), the PCR products were purified using the Promega PCR kit according to the manufacturer's instructions. Purified products were eluted in 20 µl water. Approximately 5-10 µl of purified product was digested with 5U AvaI (New England Biolabs) in a total volume of 15 µl for 1.5 hr at 37°C. Following digestion, the products were separated by electrophoresis through an 1.8 % agarose gel. Amplification of the wildtype sequence results in a 465 bp product that is not cleaved by AvaI. Amplification of the mutated (857A(G) sequence results in a 465 bp product that is cleaved by AvaI to yield products of 404 and 61 bp. Accordingly, individuals who are homozygous for the wildtype sequence exhibit a 465 bp band. Heterozygous individuals exhibit 3 bands: 465, 404, and

61 bp bands. Individuals who are homozygous for the Q286R mutation exhibit bands of 404 and 61 bp. Restriction analysis confirmed the sequencing results of all examined individuals (Figure 9).

EXAMPLE III

Genetic Screening for PPK-Associated Mutations

The foregoing findings provide the basis for screening and diagnostic assays for assessing patients for the presence of mutations in the CTSC gene related to the pathological conditions described herein. A summary of the mutations in CTSC identified as associated with PPKs are set forth in Table 1.

Table 1. Phenotype correlations with *CTSC* mutations

Mutation[a]	Predicted Effect	Location	Phenotype[b]	# of Families
199-222del	Deletion of aa 67-74	Exon 2	PLS	1
445-446insATGT	Frameshift, termination at aa 157	Exon 3	PLS	1
458C®T	Threonine 153 Isoleucine	Exon 3	PLS	2
622-623insC	Frameshift, termination at aa 223	Exon 4	PLS	1
704G®A	Tryptophan 235 Stop	Exon 5	PLS	1
748C®T	Arginine 250 Stop	Exon 5	PLS	1
815G®C	Arginine 272 Proline	Exon 6	PLS	1
856C®T	Glutamine 286 Stop	Exon 6	PLS	2
857A®G	Glutamine 286 Arginine	Exon 6	HMS	1
898G®A	Glycine 300 Serine	Exon 7	PLS	1
901G®T	Glycine 301 Valine	Exon 7	PLS	1
901G®A	Glycine 301 Serine	Exon 7	PLS	1
910T®A	Tyrosine 304 Asparagine	Exon 7	PLS, RP	1
956A®G	Glutamic Acid 319 Glycine	Exon 7	PLS	1
1015C®T	Arginine 339 Cysteine	Exon 7	PLS	2
1019A®G	Tyrosine 340 Cysteine	Exon 7	PLS	1
1028-1029delCT	Introduction of premature termination codon	Exon 7	PLS	1
1040A®G	Tyrosine 347 Cysteine	Exon 7	PPP	2
1047delA	Frameshift, termination at aa 349	Exon 7	PLS	1
1286G®A	Tryptophan 429 Stop	Exon 7	PLS	3
1360A®G	Glutamic Acid 447 Glycine	Exon 7	PLS	1

1. cDNA numbering considering the initiator Met codon as nucleotide +1.
2. Phenotype symbols: PLS, Papillon-Lefevre syndrome; PPP, Prepubertal periodontitis;
HMS, Haim Munk syndrome; RP, Retinitis pigmentosa

While the mutations described in the previous examples are associated with certain pathological conditions, it is important to note that the CTSC gene contains many polymorphisms. Many of these genetic changes are not associated with the disease state. The genetic changes assessed by the methods of the present invention must be associated with the production of an aberrant CTSC protein. Accordingly, a suitable assay for diagnosing this disorder includes the step of differentiating harmless polymorphisms from those

mutations which give rise to PPKs and periodontal disorders. These include changes in the coding sequence which give rise to decreased mRNA stability as compared to wild type CTSC mRNA. Alternatively cathepsin C enzymatic activity can be compared between altered CTSC coding sequences and nucleic acids encoding the wild type enzyme. Such assays are well known in the art and need not be set forth here. See for example, McGuire et al., Archives of Biochemistry and Biophysics 295:280-8, 1992; McDonald et al., J. of Biological Chemistry 244:2693-26709, 1969; Metroione et al, Biochemistry 5:1597-1604, 1966; and Vanha-Perttula et al., Histochemie 5:170-181, 1965.

References:

1. Papillon, M.M., and Lefevre, P. Deux cas de keratodermie palmaire et plantaire symetrique familiale (maladie de Meleda) chez le frere et la soeur. Coexistence dans les deus cas d'alterations dentaires grabes. *Bull Soc Fr Dermatol Syphilis* 1924;**31**:82-87.

2. Stevens, H.P., Kelsell, D.P., Bryant, S.P., Bishop, D.T., Spurr, N.K., Weissenbach, J., Marger, D., Marger, R.S., Leigh, I.M. Linkage of an American pedigree with palmoplantar keratoderma and malignancy (palmoplantar ectodermal dysplasia type III) to 17q24. Literature survey and proposed updated classification of the keratodermas. *Arch Dermatol*. 1996;**132**:(6) 640-51.

3. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {MIM 245000}: {8/20/98}: *World Wide Web URL*: http://www.ncbi.nlm.nih.gov/omim/

4. Gorlin, RJ, Sedano H, and Anderson VE (1964), RJ. The syndrome of palmar-plantar hyperkeratosis and premature periodontal destruction of the teeth. *J Pediatr* 1964; **65**:895-908.

5. Haneke, E. The Papillon-Lefevre syndrome: keratosis palmoplantaris with periodontopathy. Report of a case and review of the cases in the literature. *Hum Genet* 1979;**51**:(1) 1-35.

6. Laass, M.W., Hennies, H.C., Preis, S., Stevens, H.P., Jung, M., Leigh, I.M., Wienker, T.F., Reis, A..

Localization of a gene for Papillon-Lefevre syndrome to chromosome 11q14-q21 by homozygosity mapping. *Hum Genet* 1997;**101**:(3) 376-82.

7. Fischer, J., Blanchet-Bardon, C., Prud'homme, J.F., Pavek, S., Steijlen, P.M., Dubertret, L., Weissenbach, J. Mapping of Papillon-Lefevre syndrome to the chromosome 11q14 region. *Eur J Hum Genet 1997;* **5**:(3)156-60.

8. Hart, T.C., Bowden, D.W., Ghaffar, K.A., Wang, W., Cutler, C.W., Cebeci, I., Efeoglu, A., Firatli, E. Sublocalization of the Papillon-Lefevre syndrome locus on 11q14-q21. *Amer J Med Genet* 1998;**79**:(2) 134-139.

9. Hart TC, Bowden DW, Hart PS, Walker SJ, Callison SA, Bobby PL, Firatli E. An integrated physical and genetic map of the PLS locus interval on chromosome 11q14. Mammalian Genome *In Press*

10. Kirschke, H., Barrett, A.J., Rawlings, N.D. Proteinases 1: Lysosomal cysteine proteinases. In: Protein Profile **Vol.2,** P. Sheterline, ed. (London, UK: *Academic Press Ltd.)* 1995;1587-1643.

11. Rao, N.V., Rao, G.V., Hoidal, J.R. Human dipeptidyl-peptidase I. *J Biol Chem*.1997; **272**:10260-10265.

12. Paris, A., Strukelj, B., Pungercar, J., Renko, M., Dolenc, I., Turk, V. Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Lett* 1995;**369**:326-330.

13. Dolenc, I., Turk, B., Pungercic, G., Ritonja, A., Turk, V. Oligomeric structure and substrate induced inhibition of human cathepsin C. *J Biol Chem* 1995;**270**:21626-31.

14. Pham, C. T. N., Armstrong, R. J., Zimonjic, D.B., Popescu, N. C., Payan, D. G., Ley, T. J. Molecular cloning, chromosomal localization, and expression of murine dipeptidyl peptidase I. *J Biol Chem* 1997;**272**:10695-10703.

15. McGuire, M.J., Lipsky, P.E., Thiele, D.L. Cloning nad characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C). *Biochimica et Biophysica Acta* 1997; **1351**:267-73.

16. Hakeda, Y., Kumegawa, M. Osteoclasts in bone metabolism. Kaibogaku Zasshi *J Anat* 1991;**66**:215-225.

17. Hart , T.C. and Shapira, L. Papillon Lefevre syndrome. *Periodontol 2000* 1994;**6**, 88-100.

18. Firatli, E., Tuzun, B., Efeoglu, A. Papillon-Lefevre syndrome. Analysis of neutrophil chemotaxis. *J Periodontol* 1996;**67**:617-20.

19. Firatli, E., Gurel, N., Efeoglu, A., Badur, S. Clinical and immunological findings in 2 siblings with Papillon-Lefevre syndrome. *J Periodontol* 1996;**67**:1210-5.

20. Itin, P.H. Classification of autosomal dominant palmoplantar keratoderme: past-present-future. *Dermatol* 1992;**185**:163-165.

21. Haim,S.,Munk J: Keratosis palmo-plantaris congenita, with periodontosis, arachnodactyly and peculiar deformity of the terminal phalanges. *Br J Dermatol* 1965;**77**:42-45.

22. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {MIM 245010}: {5/20/97}: *World Wide Web URL*: http://www.ncbi.nlm.nih.gov/omim/

23.Gorlin, R. J., Pindborg, J. J., Cohen, M. M., Jr. Syndromes of the Head and Neck. New York: McGraw-Hill (pub.) (2nd ed.) 1976. Pp. 373-376.

24. Puliyel JM, Sridharan Iyer KS. A syndrome of keratosis palmo-plantaris congenita, pes planus, onychogryphosis, periodontosis, arachnodactyly and a peculiar acro-osteolysis. Brit J Derm 1986;115:243-248.

25. Hacham-Zadeh S, Schaap T, Cohen MM. A genetic analysis of the Papillon-Lefevre syndrome in a Jewish family from Cochin. Am J Med Genet 1978;2:153-157.

26 Gorlin RJ,Cohen MM jr., Levin LS. Syndromes of the head and neck . 3rd edn. New York. Oxford University Press, 1990:853-854 (Oxford Monographs on Medical Genetics, No.19).

27. Hart T, Hart PS, Bowden D, Walker S, Callison S, Michalec M, Firatli E. Mutations of the cathepsin C gene are responsible for Papillon Lefevre syndrome. In Press J Med Genet

28. Hart TC, Stabholz A, Meyle J, Shapira L, Van Dyke TE, Cutler CW, Soskolne WA. Genetic studies of syndromes with severe periodontitis and palmoplantar hyperkeratosis. J Periodont Res 1997;32:81-89.

29. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. A second-generation linkage map of the human genome.

Nature 1992;359:794-801.

30. Lucker GP, Van de Kerkhof PC, Steijlen PM. The hereditary palmoplantar keratoses: an updated review and classification. Brit J Derm 1994;131(1):1-14.

31. Smith P, Rosenzweig KA. Seven cases of Papillon Lefevre Syndrome. Periodontics 1967;5:42-6.

32. Cohen T, Bloch N. Immigrant Jews from Cochin. In Goldschmidt E (ed) The Genetics of Migrant and Isolate Populations. Baltimore: Williams & Wilkins 1963;352.

33. Wilkie AO. Craniosynostosis: genes and mechanisms. Hum Molec Genet 1997;6:1647-56.

34. Hola-Jamriska L, Tort JF, Dalton JP, Day SR, Fan J, Asskov J and Brindley PJ. Cathepsin C from Schistosoma japonicum; cDNA encoding the preproenzyme and its phylogenetic relationships. Eur J Biochem 1998;255:527-534.

35. Wolters PJ, Raymond WW, Blount JL, Caughey GH. Regulated expression, processing, and secretion of dog mast cell dipeptidyl peptidase I. J Biol Chem 1998;25:15514-15520.

36. Mackenzie IC, Rittman G, Gao Z, Leigh I, Lane EB. Patterns of cytokeratin expression in human gingival epithelia. J Periodont Res 1991;26468-478.

37. Hormia M, Sahlberg C, Thesleff I, Airenne T. The epithelium-tooth interface--a basal lamina rich in laminin-5 and lacking other known laminin isoforms. J Den Res (1998);77:1479-85.

38. Stabholz A, Taichman NS, Soskolne WA. Occurrence of Actinobacillus actinomycetemcomitans and anti leukotoxin antibodies in some members of an extended family affected by Papillon-Lefevre syndrome. J Periodontol 1995;66:653-57.

39. Page RC. Altman LC. Ebersole JL. Vandesteen GE. Dahlberg WH. Williams BL. Osterberg SK. Rapidly progressive periodontitis. A distinct clinical condition. Journal of Periodontology. 54(4):197-209, 1983

40. Page RC. Bowen T. Altman L. Vandesteen E. Ochs H. Mackenzie P. Osterberg S. Engel LD. Williams BL. Prepubertal periodontitis. I. Definition of clinical disease entity. Journal of Periodontology. 54(5):257-71, 1983

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.